Self-Organising Load Balancing for OFDMA Cellular Networks

Lexi Xu

Submitted for the degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary, University of London

April 2013

To my family

Declaration: I hereby declare that the work presented in this thesis is solely my work and that to the best of my knowledge the work is original except where indicated by reference to respective authors.

Lexi Xu

Date: 16/April/2013

Abstract

In this thesis, self-organising load balancing is investigated to deal with the uneven load distribution in OFDMA based cellular networks. In single-hop cellular networks, a self-organising cluster-based cooperative load balancing (CCLB) scheme is proposed to overcome the 'virtual partner' and the 'aggravating load' problems confronted in the conventional mobility load balancing schemes. Theoretical analysis and simulation results show that the proposed scheme can effectively reduce the call blocking probability, the handover failure rate, and the hot-spot cell's load.

The proposed CCLB scheme consists of two stages: partner cell selection and traffic shifting. In the partner cell selection stage, a user-vote assisted clustering algorithm is proposed, which jointly considers the users' channel condition and the surrounding cells' load. This algorithm can select appropriate neighbouring cells as partners to construct the load balancing cluster, and deal with the 'virtual partner' problem. In the traffic shifting stage, a relative load response model (RLRM) is designed. RLRM coordinates multiple hot-spot cells' shifting traffic towards their public partner, thus mitigating the 'aggravating load' problem of the public partner. Moreover, a traffic offloading optimisation algorithm is proposed to balance the hot-spot cell's load within the load balancing cluster and to minimise its partners' average call blocking probability.

The CCLB scheme is modified to apply in multi-hop cellular networks with relays deployed. Both fixed relay and mobile user relay scenarios are considered. For fixed relay cellular networks, a relay-level user shifting algorithm is proposed. This algorithm jointly considers users' channel condition and spectrum usage of fixed relay, in order to reduce the handover failure rate and deal with the 'aggravating load' problem of fixed relay. In the mobile user relay scenario, the user relaying assisted traffic shifting algorithm is proposed to improve the link quality of shifted edge users, which brings about an increase in the achievable rate of shifted edge users and decrease in the handover failure rate.

Acknowledgment

I would like to express my gratitude to all those who helped me in the past three years.

The first person I want to say "thank you" to is my supervisor Dr Yue Chen. She is a knowledgeable, charming, smart, beautiful and nice teacher. During my past research time, she gave me the strongest support and guidance for both my study and daily life. She gave me lots of freedom to study an area that interested me and she encouraged me on every small progress I have ever made. Her selflessness, warm-hearted to students and passion for the research is my treasure.

I would also like to express my appreciation to Dr John Schormans, Dr Kok Keong (Michael) Chai and Dr Chris Phillips for their helpful suggestions and comments on my research. Dr John Schormans gave me a lot of useful advice when I establish the simulation platform and modify my papers. Dr Michael Chai gave me lots of valuable suggestions on the researches of load balancing and OFDMA systems. Dr Chris Phillips gave me support and encouragement, boosting my confidence and helping me to finish the thesis well.

During the past two years, many of my QMUL friends and colleagues gave me lots of help both life and research. Thanks to Kejing, Dapeng, Hongyi, Xian, Adeel, Awais, Rehana, Yousef, Fei, Geng, Yifeng, Nan, Wenxuan, Di, Haibo, Kejiong, Peng, Dan, Yun, Dantong, Lifeng, Zhijing, Yansha and so many others.

Many thanks to my BUPT team, who assisted my research although they are far from me, including Prof. Yinghai Zhang, Prof. Weidong Wang, Dr Chaowei Wang, Xinlei, Shoufeng, Gaofeng, Yuan, Beng, Zhi, Juyi, Xianwei, Liping and many others.

In my research, there are so many people I should thank. The love they have given me is enormous and selfless. With my love and gratitude, I want to dedicate this report to all the people who have ever helped me.

Contents

List of Figures

List of Tables

List of Abbreviations

2G	Second Generation
3G	Third Generation
4G	Fourth Generation
3GPP	Third Generation Partnership Project
AF	Amplify-and-Forward
ARS	Ad-Hoc Relaying Station
BS	Base Station
CBWL	Channel Borrowing Without Locking
CCLB	Cluster-based Cooperative Load Balancing
CDMA	Code Division Multiple Access
DF	Decode-and-Forward
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FRF	Frequency Reuse Factor
GBR	Guaranteed Bit Rate
GLB	Geographic Load Balancing
GoS	Grade of Service
GSM	Global System for Mobile Communication
HA	Hybrid Assignment
HM	Handover Margin
HO	Handover
HO_{off}	Cell-Specific Handover Offset
iCAR	Integrated Cellular and Ad-Hoc Relay
ISM	Industrial, Scientific and Medical

KKT	Karush-Kuhn-Tucker
LOS	Line-Of-Sight
LTE	Long Term Evolution
LTE-Advanced	Long Term Evolution Advanced
Max-C/I	Maximum Carrier to Interference Ratio
MLB	Mobility Load Balancing
MS	Mobile Station
MSFR	Modified Soft Frequency Reuse
NLOS	Non-Line-Of-Sight
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
QoS	Quality of Service
RLRM	Relative Load Response Model
RNC	Radio Network Controller
RS	Relay Station
PRB	Physical Resource Block
RRM	Radio Resource Management
RS	Relay Station
RSRP	Reference Signal Received Power
SB	Simple Borrowing
SD	Standard Deviation
SIR	Signal to Interference Ratio
SINR	Signal to Interference plus Noise Ratio
SON	Self-Organising Networks
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TS	Time Slot

TTI	Transmission Time Interval
UMTS	Universal Mobile Telecommunications System
URTS	User Relaying assisted Traffic Shifting
WCDMA	Wideband Code Division Multiple Access
WTS	Weight of Traffic Shifting
WEC	Weight of Energy Consumption

Chapter 1 Introduction

1.1 Background

The fast emerging wireless services and users' mobility may result in many users gathering in a cell, which brings large traffic demand. Thus this cell becomes heavily loaded. Meanwhile, its neighbouring cells serve a small number of users with much less traffic demand. Hence, the mobile cellular networks suffer uneven load distribution.

The uneven load distribution impacts the network performance in the following way. On one hand, due to the limited spectrum resources, the heavily loaded cell may reject access requests of new call users; due to the large traffic demand, the heavily loaded cell may be unable to provide the required quality of service (QoS) of existing users. On the other hand, its neighbouring cells may serve few users, thus these neighbouring cells are spectrum underutilised.

In order to deal with the uneven load distribution, load balancing is widely used to redistribute load among heavily loaded cell and neighbouring cells. During the load balancing process, a heavily loaded cell can select less-loaded neighbouring cells as its partner cells to shift traffic or share spectrum resources. Load balancing can reduce the load of heavily loaded cell, thus the cell allows new call users' admission requests and is able to provide existing users with the required QoS. Therefore, load balancing can effectively improve the network performance in terms of call blocking probability, QoS and spectrum utilisation.

In 3GPP Long Term Evolution (LTE) / LTE-Advanced, mobility load balancing (MLB) is an effective method to address uneven load distribution [3GPP11b]. The basic idea of MLB is that a heavily loaded cell selects its partners and adjusts cell-specific handover offset towards partners, which enlarges the hard handover area [3GPP12]. Then some users in the heavily loaded cell are handed over to those partners. However, there are some problems in the conventional MLB schemes.

In conventional MLB schemes, neighbouring cell's load has been widely used as a criterion for partner selection. This may result in the virtual partner problem, which denotes a lightly loaded neighbouring cell, while it is far from the heavily loaded cell's edge users. After partner selection, multiple heavily loaded cells may shift traffic to a public partner. Without the coordination of multiple cells, conventional MLB schemes may result in the public partner's aggravating load problem, which denotes the public partner becomes heavily loaded after traffic shifting. In addition, the shifted user may suffer the link quality degradation problem, since it may receive the reduced signal power from the partner cell.

Relay station (RS) can extend cell coverage and enhance users' performance in cell edge area. Therefore, RS is considered as an important technology in LTE-Advanced [3GPP10d]. Since most shifted users are served by RSs in fixed relay cellular networks, conventional MLB schemes may result in the RS aggravating load problem, which denotes the RS becomes heavily loaded after traffic shifting.

The research in this thesis aims at providing a self-organising load balancing algorithm, which can deal with above problems and provide better performance compared with conventional MLB schemes, e.g., lower call blocking probability, lower handover failure rate.

1.2 Research Scope

This thesis describes the research on the self-organising load balancing in Orthogonal Frequency Division Multiple Access (OFDMA) cellular networks. The self-organising load balancing is based on handover and requires some in-depth research considerations:

- How to identify a heavily loaded cell under time-varying load; How does a heavily loaded cell select its partner cells under dynamic user distribution;
- How to avoid the partners being heavily loaded, especially under the scenario of

multiple cells shifting traffic to one partner (called the public partner);

- How to automatically adjust MLB parameters, e.g., cell-specific handover offset;
- How to shift traffic in fixed relay cellular networks;
- How to overcome the link quality degradation for shifted users.

In single-hop cellular networks, a self-organising cluster-based cooperative load balancing (CCLB) scheme is proposed. The CCLB scheme is modified to apply in fixed relay cellular networks. Furthermore, this thesis investigates how to employ non-active users as mobile relay to forward the transmission data to shifted users.

1.3 Research Contribution

The major contribution of the research work is a self-organising cluster-based cooperative load balancing (CCLB) scheme. The CCLB scheme can deal with the problems confronted in conventional MLB, including virtual partner problem, aggravating load problem of public partner, aggravating load problem of RS, and the link quality degradation of shifted user.

1. CCLB scheme in single-hop cellular networks

In single-hop cellular networks, the CCLB scheme consists of two stages: partner selection and traffic shifting. In the partner selection stage, a user-vote assisted clustering algorithm is proposed. Based on users' channel condition and neighbouring cells' load, the cluster head (which denotes a heavily loaded cell) selects appropriate neighbouring cells as partners, in order to construct load balancing cluster for traffic shifting. This clustering algorithm can deal with the virtual partner problem and select suitable partners.

In the traffic shifting stage, a cell-level cooperative traffic shifting algorithm is designed. In this algorithm, the cluster head effectively shifts traffic with the cooperation of partners. This algorithm includes two steps:

- In the inter-cluster cooperation step, a relative load response model (RLRM) is proposed, which coordinates multiple cluster heads' traffic shifting requests to one public partner. RLRM can address the aggravating load problem of public partners.

- In the intra-cluster cooperation step, a traffic offloading optimisation algorithm is designed via Lagrange multipliers method. This proposed algorithm can shift the cluster head's traffic and also minimise its partners' average call blocking probability.

2. CCLB scheme in fixed relay cellular networks

The proposed CCLB scheme is modified to apply in fixed relay cellular networks. The CCLB scheme is composed of three algorithms: user-vote assisted clustering algorithm, cell-level cooperative traffic shifting algorithm, relay-level user shifting algorithm.

The user-vote assisted clustering algorithm and the cell-level cooperative traffic shifting algorithm are modified according to the features of load balancing in fixed relay cellular networks. The modified algorithms can effectively select partners and address the aggravating load problem of public partner in fixed relay cellular networks.

After the above two stages, a novel relay-level user shifting algorithm is proposed for fixed relay cellular networks. This algorithm considers users' channel condition and analyses the spectrum resources usage of RSs, in order to shift appropriate users from the cluster head to the partner's RSs. This algorithm can reduce the handover failure rate and deal with the aggravating load problem of RS.

3. User relaying assisted traffic shifting

User relaying assisted traffic shifting (URTS) algorithm is the extension of CCLB scheme. The URTS algorithm works on the stage when the cluster head shifts edge user to partner cells.

Compared with the signal power from the cluster head, the shifted user receives lower signal power from the partner cell, and hence the shifted user may suffer the link quality degradation. This thesis considers a user relaying model: a non-active user is employed as a mobile relay to forward signal for a shifted user. Based on this model, a URTS algorithm is designed. This algorithm can select suitable relay user to effectively enhance the shifted user's link quality, and keep low cost of the relay user's energy consumption.

From the introduction above, the proposed CCLB scheme is self-organising. The user-vote assisted clustering algorithm is via cell to cell communication and is adaptive to the time-

varying channel condition and load distribution. In the cell-level cooperative traffic shifting algorithm, the cluster head and partners self-optimised offload traffic within the load balancing cluster, which is also an essential feature of self-organising networks [E^{3}08].

1.4 Author's Publications

[Xu-1] Lexi Xu, Yue Chen, "Priority-based Resource Allocation to Guarantee Handover and Mitigate Interference for OFDMA Systems," IEEE PIMRC2009, 13-19 September 2009, Tokyo, Japan, pp. 783-787

[Xu-2] Lexi Xu, Yue Chen, Yue Gao, "Self-organizing Load Balancing for Relay Based Cellular Networks," IEEE CIT2010 Workshop, 29 June - 1 July 2010, Bradford, UK, pp. 791-796

[Xu-3] Lexi Xu, Yue Chen, John Schormans, Laurie Cuthbert, Tiankui Zhang, "User-Vote Assisted Self-organizing Load Balancing for OFDMA Cellular Systems," IEEE PIMRC2011, 11-14 September 2011, Toronto, Canada, pp. 217-221

[Xu-4] Lexi Xu, Yue Chen, Yue Gao, Laurie Cuthbert, "A Self-optimizing Load Balancing Scheme for Fixed Relay Cellular Networks," IET ICCTA2011, 12-14 October 2011, Beijing, China, pp. 306-311

[Xu-5] Lexi Xu, Yue Chen, Kok Keong Chai, Tiankui Zhang, John Schormans, Laurie Cuthbert， "Cooperative Load Balancing for OFDMA Cellular Networks," European Wireless, 17-20 April 2012, Poznan, Poland, pp. 1-7

[Xu-6] Lexi Xu, Yue Chen, Kok Keong Chai, John Schormans, Laurie Cuthbert, "Self-Organising Cluster-based Cooperative Load Balancing in OFDMA Cellular Networks," Wiley Wireless Communications and Mobile Computing (under minor revision)

[Xu-7] Lexi Xu, Yue Chen, KoK Keong Chai, Dantong Liu, Shaoshi Yang, John Schormans, "User Relay assisted Traffic Shifting in LTE-Advanced Systems," IEEE VTC-Spring 2013, 2-5 June 2013, Dresden, Germany (accepted)

1.5 Thesis Organisation

The rest of this thesis is organised as follows.

Chapter 2 introduces the background, including the basic concept of load balancing, and the evolution of load balancing schemes from 2G networks to LTE-Advanced networks. Then this chapter discusses MLB, including its basic idea, conventional MLB schemes, and problems faced by conventional MLB schemes. Finally, load balancing in multi-hop OFDMA cellular networks is discussed.

Chapter 3 discusses the system model and simulation platform. The simulation platform consists of three system models, including single-hop cellular networks, user relay based two-hop cellular networks, and fixed relay based two-hop cellular networks. The flowchart and key modules in this simulation platform are also introduced.

Chapter 4 begins with the problem formulation of conventional MLB schemes in single-hop OFDMA cellular networks. Then a CCLB scheme is proposed to address these problems. Specifically, a user-vote assisted clustering algorithm is designed for effective partner selection. A relative load response model is designed to address the aggravating load problem of public partner. Furthermore, the traffic offloading optimisation is proposed to minimise partners' average blocking probability. The CCLB scheme is evaluated by simulation.

Chapter 5 researches the load balancing in multi-hop networks, including fixed relay cellular networks and mobile relay cellular networks. In Section 5.2, CCLB scheme is applied in fixed relay cellular networks. The user-vote assisted clustering algorithm and the cell-level cooperative traffic shifting algorithm are modified to select partners and to mitigate the public partner's aggravating load problem. Then, a novel relay-level user shifting algorithm is proposed to deal with the RS aggravating load problem. In Section 5.3, due to load balancing, shifted user's link quality degradation problem is discussed. Then, a novel user relaying assisted traffic shifting algorithm is designed to enhance the shifted user's link quality.

Chapter 6 concludes the work in this thesis, and the direction of future work is discussed.

Chapter 2 Load Balancing in Cellular Networks

2.1 Basic Concept of Load Balancing

2.1.1 Load Balancing Scenario and Objectives

Due to the random distribution of users and the exponentially growing demand for wireless data services, mobile cellular networks face the challenges brought by the uneven load distribution. Figure 2.1 exemplifies a scenario of uneven load distribution.

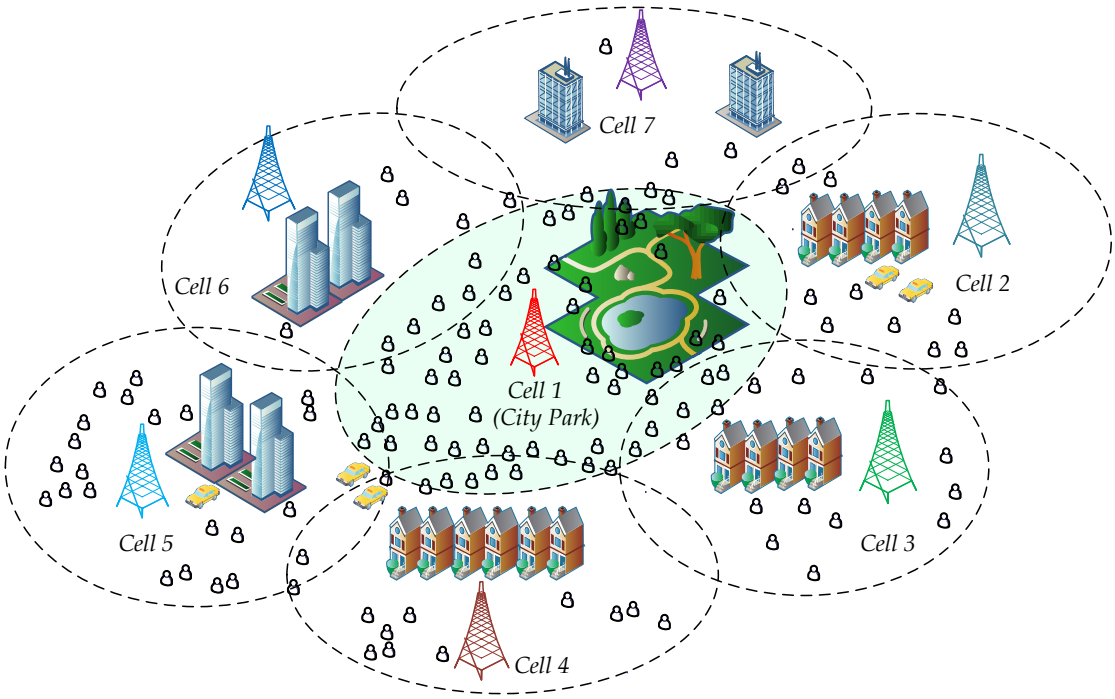


Figure 2.1 A scenario of uneven load distribution

As shown in Figure 2.1, BS_1 of $Cell_1$ is located in the city park, and surrounded by six neighbouring cells. The networks operator sets BSs coverage and pre-allocates spectrum resources according to their usual traffic load, which are estimated based on population density. Under a social event, e.g., Christmas party, music concert, thousands of users gather in the city park. The traffic generated by these users is near or even higher than the capacity of BS_1. Thus $Cell_1$ becomes heavily loaded. Meanwhile, the traffic generated by users in neighbouring cells is very low. Therefore, the cellular networks suffer an uneven load distribution.

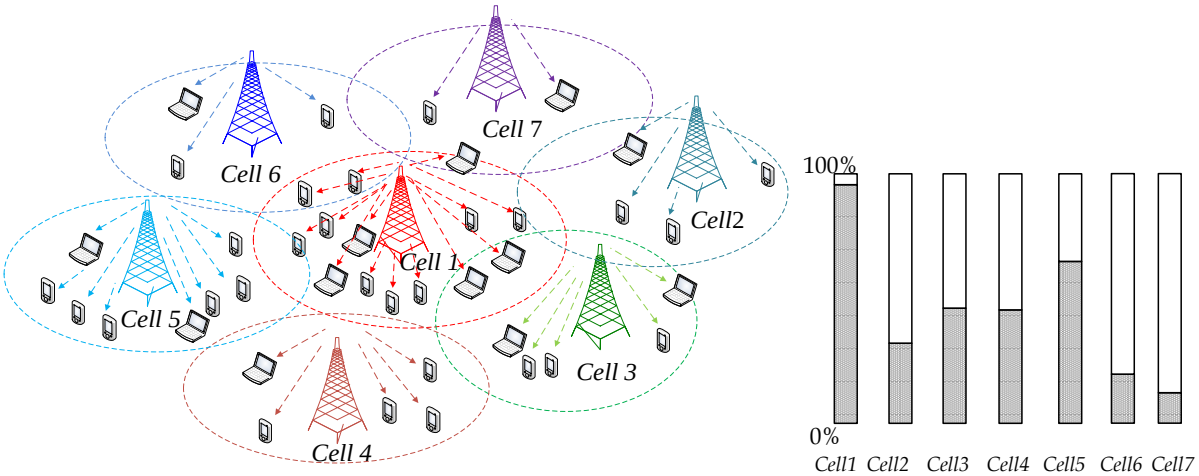
The impact of the uneven load distribution is reflected in different aspects. The heavily loaded cell, e.g., $Cell_1$ in the given scenario, may reject access requests of new call users; $Cell_1$ may not be able to provide serving users with their required QoS [KGSABS09]; the spectrum of neighbouring lighted loaded cells, such as $Cell_6$ and $Cell_7$, are underutilised.

Load balancing is one of radio resource management functionalities. Load balancing can mitigate the negative impact of the uneven load distribution, and improve the networks performance. Commonly used performance indicators for evaluating a load balancing scheme are:

- Call blocking probability
- Handover failure rate (for traffic shifting based load balancing schemes)
- Load reduction in heavily loaded cell

2.1.2 Load Balancing Process

The general process of load balancing includes three steps. The network controller, e.g., mobile switching center in 2G networks, identifies a hot-spot cell according to the cell's load condition. Then, the hot-spot cell selects less-loaded neighbouring cells as its partners. After the partner selection, the hot-spot cell takes a specific load balancing scheme to balance load with partners, e.g., borrowing idle channels from partners or shifting users at the cell edge to its partners. Figure 2.2 gives an example of uneven load distribution.



(a) Users distribution (b) Load of each cell

Figure 2.2 An example of uneven load distribution among cells

1) *Load Indicator*

Load estimation is an essential step for load balancing schemes. In this thesis, the cell load is formulated as [3GPP10b]

$$\textit{Cell Load: } L = \frac{\textit{Number of carriers used in the cell}}{\textit{Total number of carriers available to the cell}} \tag{2.1}$$

According to (2.1), a cell's load is in the range from 0% to 100% ($0\% \le L \le 100\%$). In this thesis, any cell's load can be divided into different levels:

- *Overload*: The traffic generated by users is equal to or larger than the cell capacity. Namely, all carriers are used in the cell.
- *Heavy load*: A large number of carriers are used in the cell. Namely, $100\% > L \ge L_{HL}$. L_{HL} is the threshold to identify a heavily loaded cell. In this thesis, the value of L_{HL} is 70% [SOCRATES10].
- *Light load*: A small number of carriers are used in the cell. Namely, $L < L_{HL}$.

Note that Equation (2.1) is an example of calculation cell's load [3GPP10b]. The formula to calculate cell's load is based on the specific multiple access technology and the load balancing scheme. In addition, different load balancing schemes may have different

methods or values of threshold to identify cell's load level.

Load balancing can be triggered when the load of a cell is equal to or higher than the heavily loaded threshold L_{HL}. In this thesis, a cell with a load above L_{HL} is defined as a hot-spot cell. For example, in Figure 2.2, $Cell_1$ is a hot-spot cell and triggers load balancing.

2) *Partner Selection*

The second step of balancing $Cell_1$ load is to select one or more neighbouring cells (e.g., $Cell_2$, $Cell_3$…$Cell_7$ in Figure 2.2) as the partners, which are also called as target cells or selected neighbouring cells in some conventional load balancing schemes. If $Cell_1$ selects an inappropriate neighbouring cell as its partner, such as $Cell_5$, $Cell_5$ may become heavily loaded after load balancing. Then, both hot-spot $Cell_1$ and partner $Cell_5$ will suffer high load balancing handover failure rate and high call blocking probability. In many conventional load balancing schemes, such as [ZY89] [TY03] [NA07] [KAPTK10] [WTJLHL10] [YLCW12], this step is based on neighbouring cell's load.

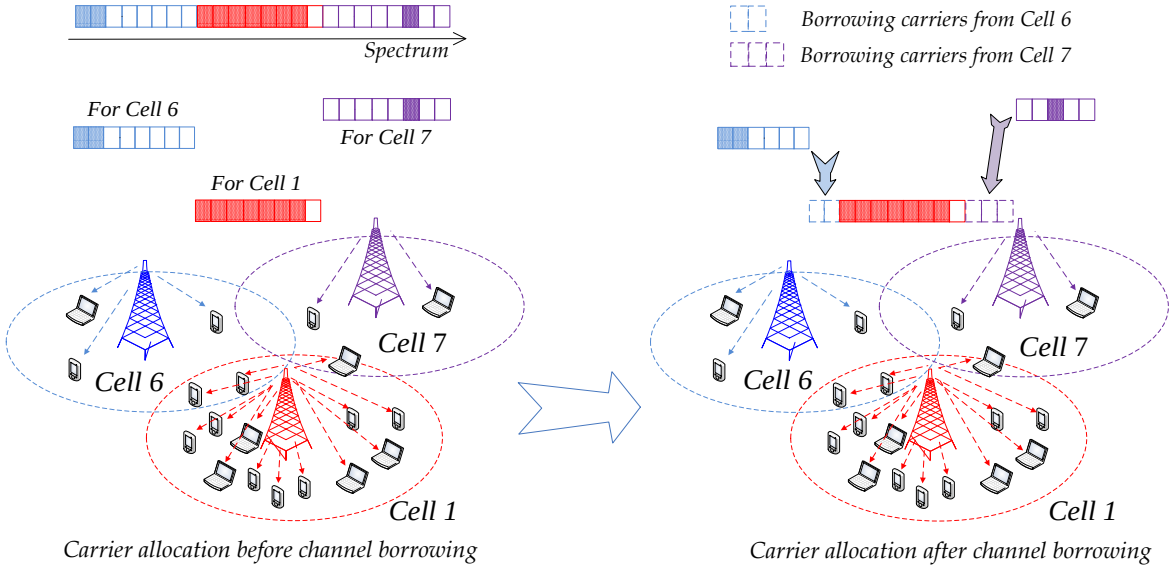
3) Channel Borrowing or Traffic Shifting

After selecting partner/s, the hot-spot cell employs a specific load balancing scheme to balance the load between the hot-spot cell and selected partners. In general, conventional load balancing schemes can be divided into two categories: *channel borrowing* schemes shown in Figure 2.3(a), and *traffic shifting* schemes shown in Figure 2.3(b).

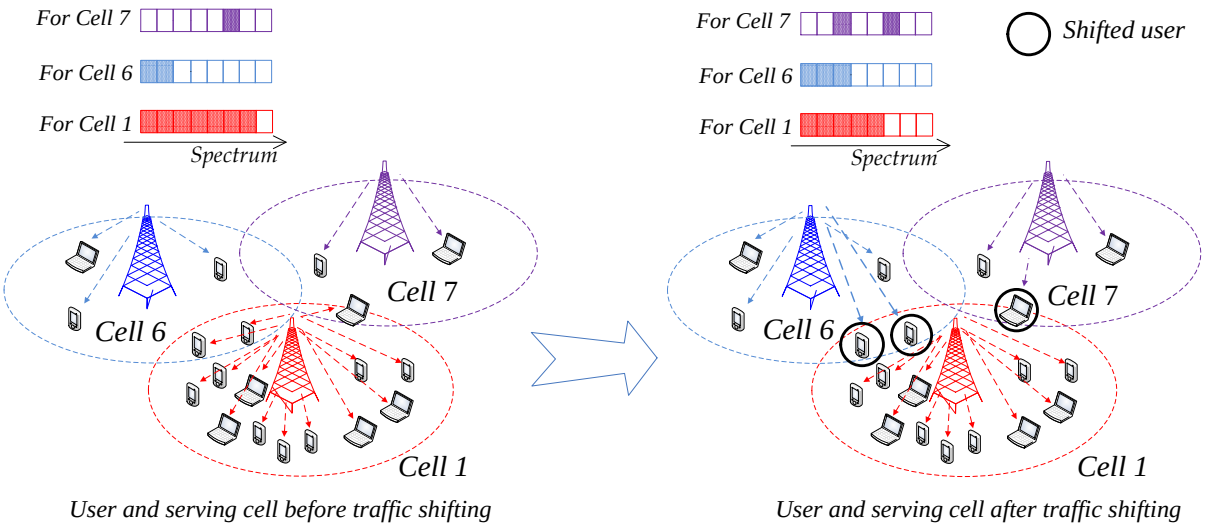
The basic idea of channel borrowing schemes is that the hot-spot cell borrows idle spectrum from partner cells. Channel borrowing schemes are suitable for cellular networks with frequency reuse factor (FRF) greater than 1, where neighbouring cells use different frequency spectrum to mitigate inter-cell interference. As shown in Figure 2.3(a), $Cell_1$, $Cell_6$ and $Cell_7$ use non-overlapping spectrum to serve users in their coverage. When $Cell_1$ is heavily loaded, $Cell_1$ can borrow idle carriers from Cell_6 and Cell_7 to increase the available carriers. Then $Cell_1$ can serve more new call users and provide better QoS for existing users.

Cellular networks with FRF>1 can effectively mitigate inter-cell interference. However, the networks with FRF>1 has lower overall spectrum efficiency than networks that employ the full frequency reuse (FRF=1), where different cells use the overlapping spectrum [RY10]. In this type of networks, load balancing is based on *traffic shifting*: the hot-spot cell shifts edge

users to its partner cells via handover (a user is defined as cell edge user if the user's reference signal received power (RSRP) difference between two neighbouring cells is lower than a threshold, e.g., 3dB [FSCK10] [SKMNT10]). Figure 2.3(b) shows a cellular network with FRF=1, where $Cell_1$, $Cell_6$ and $Cell_7$ use the identical spectrum bandwidth. When $Cell_1$ becomes a hot-spot, it shifts some of edge users to $Cell_6$ and $Cell_7$. This reduces the number of users served by $Cell_1$. The released carriers/spectrum from the shifted users can be allocated to new call users or exiting users in $Cell_1$.



(a) Channel borrowing based load balancing



(b) Traffic shifting based load balancing

Figure 2.3 Channel borrowing and traffic shifting based load balancing schemes

In the past two decades, load balancing has been investigated in both academia and industry. The specific load balancing scheme is related to specific multiple access technology and

frequency reuse technology in cellular networks. Figure 2.4 illustrates the widely used load balancing schemes from the 2G to the future LTE/LTE-Advanced cellular networks.

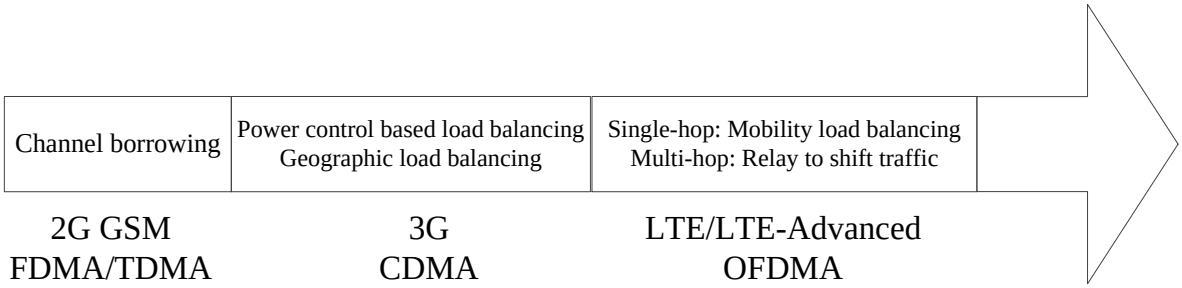


Figure 2.4 Typical load balancing schemes for different generations of cellular networks

2.2 Load Balancing in 2G GSM Networks

2.2.1 Multiple Access and Frequency Reuse

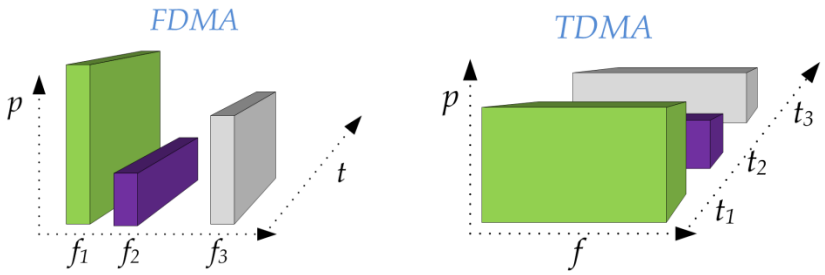


Figure 2.5 Multiple access: FDMA and TDMA [Chen03]

Figure 2.5 shows the two multiple access technologies used in Global System for Mobile Communication (GSM) networks. In Frequency Division Multiple Access (FDMA), signals for different users are transmitted in different frequency bands at the same time. In Time Division Multiple Access (TDMA), signals for different users are transmitted in the same frequency band at different times [Chen03].

GSM networks jointly employ FDMA and TDMA. In FDMA, the network operator divides the whole spectrum into several carriers, and each carrier has a unique frequency. In TDMA, each carrier is divided into eight time slots. Therefore, users transmit their signals at different time slots of different carriers.

If neighbouring cells assign the same time slot of co-channel carriers to their users, these users will suffer severe co-channel interference. In order to deal with this problem, GSM

network operators employ the frequency reuse technology to segregate co-channel carriers in neighbouring cells. Figure 2.6 shows the typical 7-cell frequency reuse technology (FRF=7). The cellular networks consist of three clusters, and each cluster includes 7 cells. The total carriers are divided into 7 groups of carriers, as *group A, B, C, D, E, F, G*, respectively. In a cluster, each cell is pre-allocated one corresponding group of carriers. In order to mitigate co-channel interference, a group of carriers can be reused in neighbouring clusters' cells, if the distance is longer than the *minimum frequency reuse distance*.

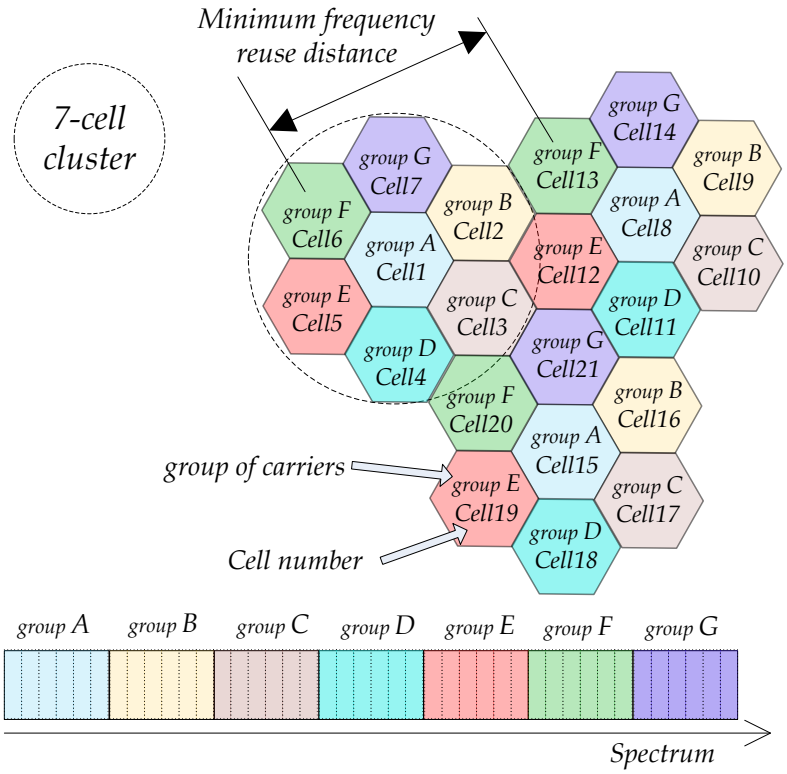


Figure 2.6 7-cell frequency reuse technology in GSM

2.2.2 Load Balancing Schemes

Channel borrowing is a popular load balancing method in GSM networks [EP73] [JR93a] [JR93b]. The basic idea is that the hot-spot cell borrows idle carriers from intra-cluster neighbouring cells. For example, in Figure 2.6, it is assumed that $Cell_1$ is a hot-spot and uses all carriers in *group A*, while $Cell_6$ and $Cell_7$ are lightly loaded. Then $Cell_1$ borrows part of idle carriers from intra-cluster neighbouring cells, including idle carriers in *group F* from $Cell_6$, and idle carriers in *group G* from $Cell_7$.

This section introduces three typical channel borrowing schemes.

- Simple borrowing scheme (SB): A hot-spot cell borrows the idle carriers from intra-cluster neighbouring cells, and the channel locking mechanism is used [ZY89].

  Channel locking mechanism: This mechanism aims at reducing the co-channel interference resulting from carriers borrowing. Hence, when the hot-spot cell borrows a carrier, neighbouring clusters' cells within the *minimum frequency reuse distance* cannot use this carrier [JR94]. For example, in Figure 2.6, if $Cell_1$ borrows a carrier (in *group F*) from $Cell_6$, neighbouring clusters' $Cell_{13}$ and $Cell_{20}$ cannot use this carrier (in *group F*) as well. This is because that the co-channel distance of $Cell_1$-*to*-$Cell_{13}$ and that of $Cell_1$-*to*-$Cell_{20}$ are shorter than the *minimum frequency reuse distance*.

- Hybrid assignment scheme (HA): HA is also based on channel borrowing [KG78] [ZY89]. In HA, each cell divides its carriers into two subsets: one subset carriers can only be used by the original cell; while the other subset carriers can be borrowed under the channel locking mechanism, in order to mitigate the co-channel interference [ZY89].

- Channel borrowing without locking scheme (CBWL): Because the channel locking mechanism mitigates the co-channel interference with the expense of the low spectrum utilisation of neighbouring clusters, CBWL is designed in [JR93b] [JR94]. In CBWL, the hot-spot cell allocates the borrowed carriers to users in cell inner area. Then, the hot-spot cell transmits signals to these users with reduced transmit power. Thus, the co-channel interference yielded by carriers borrowing is slightly heavier, compared with that in SB/HA. Therefore, channel locking mechanism is not necessary in CBWL, and the co-channel carriers can also be used by cells in neighbouring clusters [JR94]. CBWL can achieve a more effective spectrum utilisation than SB and HA.

2.3 Load Balancing in 3G CDMA Networks

2.3.1 Multiple Access

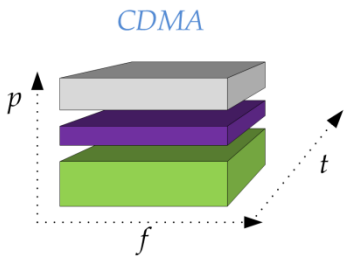


Figure 2.7 Multiple access: CDMA [Chen03]

3G standards (UMTS, cdma2000) employ wideband Code Division Multiple Access (CDMA) technology as shown in Figure 2.7. Signals for different users are identified via spreading code in spread-spectrum [Goldsmith05]. In a cell, users transmit signals in the same frequency band at the same time. Therefore, a user's signal acts as the interference to other users [NADN06]. Load balancing can improve CDMA networks performance through reducing the number of users in the hot-spot cell, thus mitigating the intra-cell interference.

2.3.2 Load Balancing Schemes

In 3G CDMA networks, all cells use the same spectrum. This leaves little space for channel borrowing. Therefore, load balancing takes a different approach: the hot-spot cell shifts some of serving traffic to less-loaded neighbouring cells [NPPDBC03] [WZ05] [NADN06] [Yao07]. In this thesis, two types of load balancing in CDMA networks are discussed, including power control based load balancing scheme and geographic load balancing scheme.

1) *Power Control based Load Balancing*

Figure 2.8 illustrates the basic idea of the power control based load balancing scheme in WCDMA networks. The hot-spot BS_1 reduces its channel transmit power or rejects edge users' requests of increasing transmit power [WZ05]. These mechanisms could shift some users to the lightly loaded BS_2. The reduced number of users in $Cell_1$ could improve the SIR (signal to interference ratio) of $Cell_1$'s serving users.

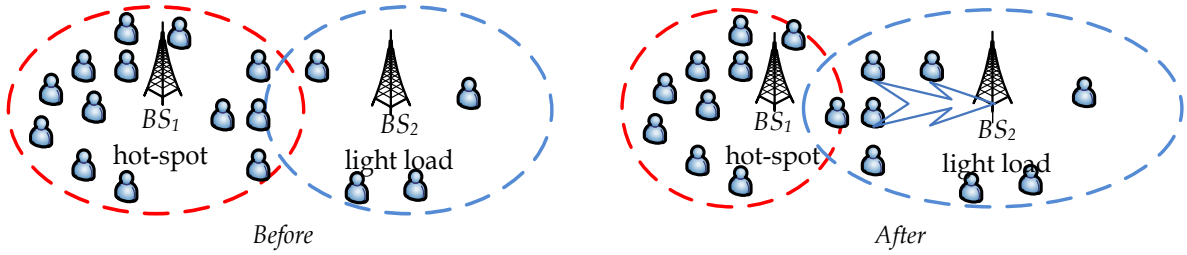


Figure 2.8 Power control based load balancing

2) *Geographic Load Balancing*

In previous research at Queen Mary University of London, geographic load balancing (GLB) is researched in [NPPDBC03] [Yao07]. The pre-condition of GLB is that each BS equips smart

antennas. Smart antennas employ smart signal processing algorithms to identify the signal direction of arrival, and then track the antenna beam of the target user dynamically [Yao07].

More specifically, the radio network controller (RNC) collects the users' location information in order to know the time-varying traffic distribution in cellular networks. Then, RNC uses sophisticated computation, such as genetic algorithm, to optimise each cell's coverage. Finally, RNC adjusts the smart antennas' pattern. In this way, GLB intelligently changes the cellular coverage according to the time-varying geographic traffic distribution.

Compared with the power control based load balancing scheme, GLB can adjust cell coverage more accurately. This is because that the sophisticated computation can adjust smart antenna pattern precisely, thus achieving good traffic distribution in cellular networks. The limitation of GLB is that BSs need to be equipped with smart antennas, which are more expensive than the ordinary three-sector antennas.

2.4 Load Balancing in Single-Hop OFDMA Networks

Compared with 3G networks, 4G networks, such as 3GPP LTE/LTE-Advanced networks, put forward higher data rate requirements of services. The CDMA technique becomes the bottle-neck for developing higher-speed mobile networks. Due to the high spectrum efficiency of Orthogonal Frequency Division Multiplexing (OFDM), LTE/LTE-Advanced networks employ Orthogonal Frequency Division Multiple Access (OFDMA) as the multiple access technology.

2.4.1 OFDM/OFDMA

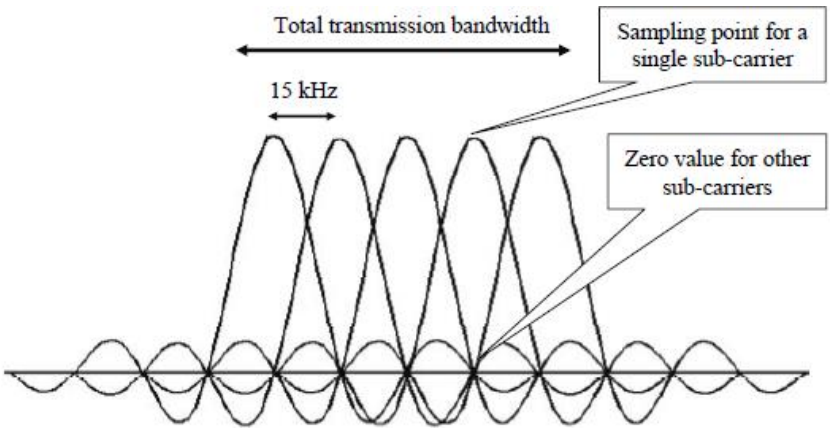


Figure 2.9 OFDM subcarriers [HT09]

The basic idea of an OFDM system is to use narrow, mutually orthogonal subcarriers to carry data. As shown in Figure 2.9 (Figure 4.4 in [HT09]), OFDM divides the high rate data stream into several parallel, low rate data streams. Each low rate data stream is assigned to one subcarrier for transmission. At the sampling instant of a single subcarrier, the other subcarriers have a zero value. Therefore, the subcarriers are orthogonal.

OFDMA is a multiple access method of the OFDM technology. OFDMA is achieved by assigning different subcarriers to different users.

The OFDM/OFDMA brings following benefits for cellular networks:

High spectral efficiency: A cell allocates different subcarriers to different users. Due to the orthogonality of subcarriers, the intra-cell interference is mitigated significantly. Therefore, OFDMA system can achieve a high rate [Xiao10].

Anti-fading: OFDMA system can effectively combat frequency-selective fading. It is due to the fact that OFDM divides the wideband transmission into narrowband transmission on several subcarriers, each subcarrier can be employed as a flat fading channel [HT09].

Flexible resource allocation: OFDMA system can select certain subcarriers for transmission according to the channel condition, thus achieving flexible resource allocation; OFDMA system can also fully make use of frequency diversity and multi-user diversity to achieve good system performance [HT09] [Xiao10].

2.4.2 Mobility Load Balancing

In order to achieve high cell capacity, one of frequency reuse technologies considered in LTE/LTE-Advanced networks is that all cells share the same spectrum (FRF=1) [3GPP10c] [RY10]. Channel borrowing based load balancing schemes, which are widely used in 2G GSM networks, are not widely used in these networks.

In OFDMA based LTE/LTE-Advanced cellular networks, the intra-cell interference is negligible due to the orthogonality of subcarriers. Besides, LTE/LTE-Advanced networks are distributed control. Power control based load balancing schemes may bring the problems of coverage hole and signalling overhead in distributed control networks. Therefore, power control based load balancing schemes, which are used in 3G CDMA networks, are not widely used in OFDMA based LTE/LTE-Advanced networks.

In order to effectively balance the load in LTE/LTE-Advanced networks, 3GPP release-8 defines mobility load balancing (MLB) as a SON[1] (self-organising networks) functionality [3GPP08b]. MLB aims at shifting the traffic load from a hot-spot cell to less-loaded neighbouring cells, via adjusting the cell-specific handover offset (HO_{off}) to enforce handover.

Generally, MLB is composed of two stages: partner selection and traffic shifting. In the partner selection stage, the hot-spot cell selects less-loaded neighbouring cells as partners, which are also called as target cells or selected neighbouring cells in some MLB schemes. This stage in many conventional MLB schemes is based on neighbouring cell's load. In the traffic shifting stage, the hot-spot cell calculates the amount of shifting traffic and adjusts HO_{off} towards each partner. The adjusted HO_{off} enlarges the handover area, thus shifting cell edge users to selected partner cells. The traffic shifting stage is illustrated in Figure 2.10, where $Cell_1$ is a hot-spot and intends to offload traffic to partner $Cell_2$. However, the user's $RSRP_2$ from BS_2 is weaker than $RSRP_1$ from BS_1, and hence the edge user is unable to trigger handover. In order to shift this edge user, BS_1 adjusts its HO_{off} towards BS_2. Once the hard handover condition (Event A3 in [3GPP12]), which is shown in (2.2), is meet, the user will be handed over to BS_2.

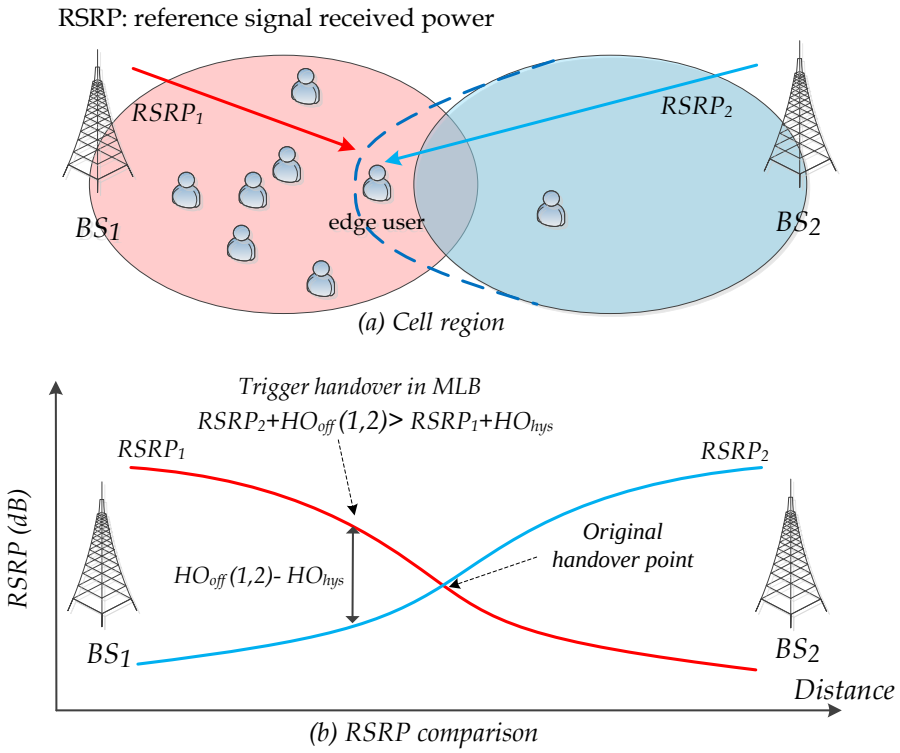


Figure 2.10 Illustration of traffic shifting stage in MLB

[1]SON aims at adapting to time-varying network environment with low maintain costs [NGMN07].

$$HO_{off}(1, 2) + RSRP_2 > RSRP_1 + HO_{hys} \quad (2.2)$$

where HO_{hys} is the handover hysteresis, HO_{hys} is fixed and HO_{hys}=2*dB* [LGK10].

HO_{hys} can ensure that $HO_{off}(1,2)+RSRP_2$ is *2dB* higher than $RSRP_1$, in order to deal with the ping-pong handover. The ping-pong handover denotes that the user is handed over to $Cell_2$ and then it is handed over back to $Cell_1$ [JBTMK10].

2.4.3 Mobility Load Balancing Schemes Introduction

In recent years, MLB has drawn a lot of attention from both industry and academia. This section introduces the conventional MLB schemes in [NA07] [KAPTK10] [LLZL10] [LSJB10] [ZQMZ10a].

2.4.3.1 Handover Adaptation for MLB

Among early research of MLB, the contribution of Ridha Nasri and Zwi Altman in [NA07] is the milestone. Ridha Nasri *et al.* proposed a general principle of cell-specific handover offset (HO_{off}) adjustments ([NA07] employed handover margin (HM) in the handover condition determination, instead of HO_{off}. Assuming $HO_{off}(i,j)$ is the handover offset from $Cell_i$ to $Cell_j$, and $HM(i,j)$ is the handover margin from $Cell_i$ to $Cell_j$, their values follow $HO(i, j) = -1 \times HM(i, j)$. In order to keep consistence in this thesis, Section 2.4.3.1 uses HO_{off}, which is in employed by 3GPP [3GPP12], to introduce the general principle).

In [NA07], the hot-spot cell chooses all lightly loaded neighbouring cells as its partners, via comparing cells load. Since the traffic direction is from a hot-spot cell to each partner, based on the load difference between $Cell_i$ and $Cell_j$, the hot-spot $Cell_i$ adjusts $HO_{off}(i,j)$ to partner $Cell_j$ via the $HO_{off}(i,j)$ adjustment function $f(L_i, L_j)$ in (2.3).

$$HO_{off}(i, j) = f(L_i, L_j) \quad (2.3)$$

where L_i and L_j are the load of $Cell_i$ and $Cell_j$, respectively. $0\% \le L_i, L_j \le 100\%$. $L_i - L_j$ is the load difference between $Cell_i$ and $Cell_j$, and $-1 \le L_i - L_j \le 1$.

In [NA07], Ridha Nasri and Zwi Altman proved that the $HO_{off}(i,j)$ adjustment function

$f(L_i, L_j)$ should satisfy the following general principle:

(a) $f(L_i, L_j)$ is an increasing function of $L_i - L_j$, under $-1 \le L_i - L_j \le 1$. When $L_i - L_j = -1$, the value of $f(L_i, L_j)$ equals the minimum handover offset HO_{off}^{min}. When $L_i - L_j = 1$, the value of $f(L_i, L_j)$ equals the maximum handover offset HO_{off}^{max}.

(b) $f(L_i, L_j) + f(L_j, L_i) = f(0)$.

where $f(L_i, L_j)$ is the $HO_{off}(i,j)$ adjustment function from $Cell_i$ to $Cell_j$, $f(L_j, L_i)$ is the $HO_{off}(j,i)$ adjustment function from $Cell_j$ to $Cell_i$. $f(0)$ is the value of the planned handover offset when the uniformity of cell loads ($L_i - L_j = 0$) are reached [NA07].

To our knowledge, a typical value of $f(0)$ is 0, and the minimum handover offset $HO_{off}^{min} = -HO_{off}^{max}$. Then, we illustrate the general principle in Figure 2.11. (a) keeps the value of $HO_{off}(i,j)$ is in the range from HO_{off}^{min} to HO_{off}^{max}. (b) is to mitigate a shifted user (from $Cell_i$ to $Cell_j$) handing over back to the hot-spot cell (from $Cell_j$ to $Cell_i$).

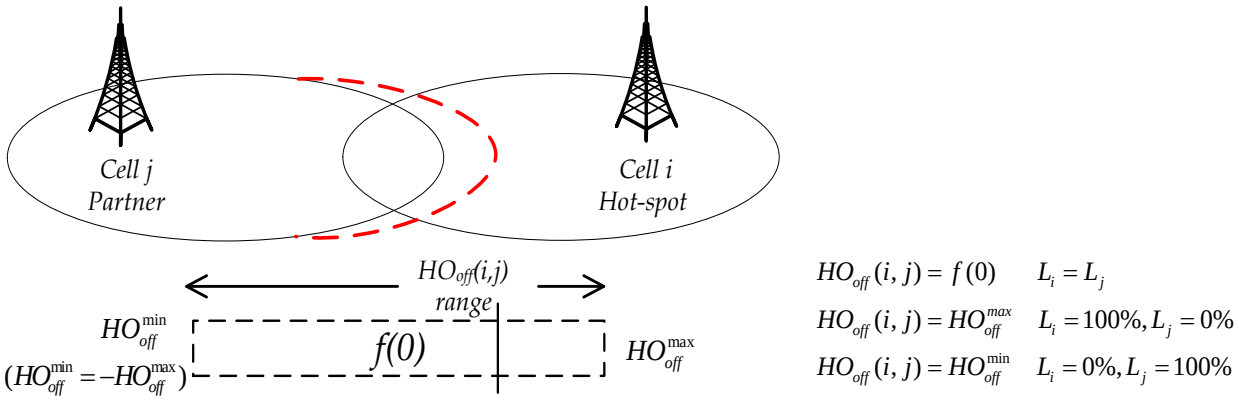


Figure 2.11 Illustration of HO_{off} adjustment in [NA07]

2.4.3.2 Precise HO_{off} adjustment based MLB

In order to precisely shift traffic to balance load among cells, based on the general principle of HO_{off} adjustment in [NA07], Raymond Kwan *et al.* researched the *precise HO_{off} adjustments* in [KAPTK10].

In [KAPTK10], a hot-spot cell selects all lightly loaded neighbouring cells, whose load are ρ_{th} lower than the hot-spot cell's load, as its partners. Then, the hot-spot $Cell_i$ precisely adjusts $HO_{off}(i,j)$ towards $Cell_j$, in order to equalise the load between the two cells. The $HO_{off}(i,j)$ adjustment process is illustrated in Figure 2.12. $Cell_i$ gradually adjusts $HO_{off}(i,j)$ with the offset step size ζ ($HO_{off}(i,j) \leftarrow HO_{off}(i,j)+\zeta$, and $\zeta=0.5dB$), until the two cells' load difference meet the requirement of traffic shifting in [KAPTK10].

Similarly, Weihao Lv *et al.* [LLZL10] also designed a precise HO_{off} adjustment based MLB scheme.

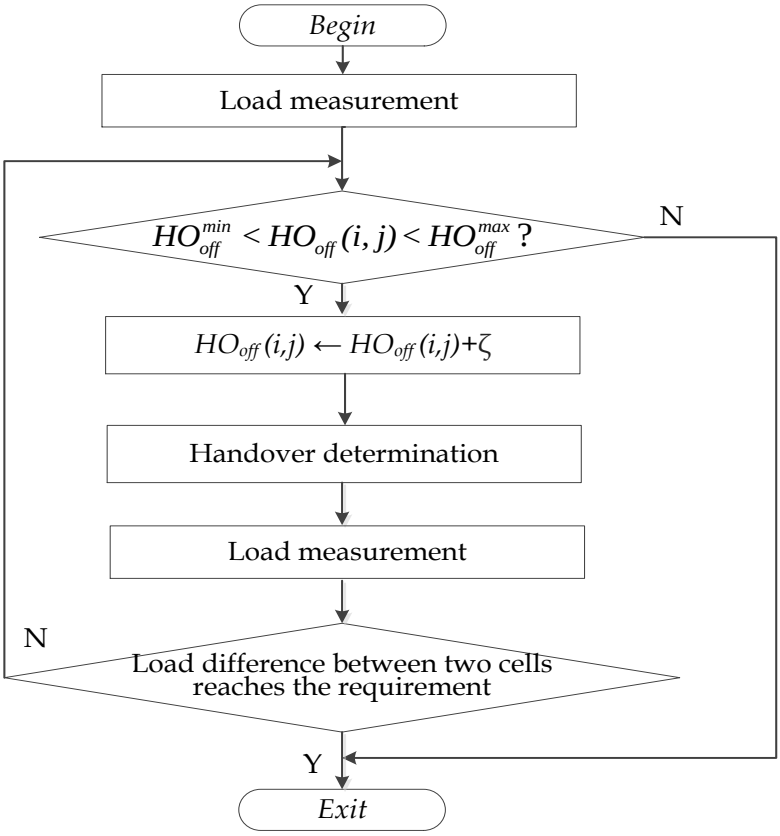


Figure 2.12 Illustration of HO_{off} adjustment process in [KAPTK10]

2.4.3.3 Load Increment Estimation based MLB

The preceding MLB schemes, such as [NA07], [KAPTK10] and [LLZL10], mainly consider the load reduction of the hot-spot cell. Due to the limited spectrum resources of the partner, the partner may have no capability to serve all shifted users and even reject the handover requests.

Subsequently, Andreas Lobinger *et al.* [LSJB10] proposed a *load increment estimation based*

MLB scheme. After partner selection (partner cell is called target eNB in [LSJB10]), the hot-spot cell estimates each partner cell's load increment, which is resulted from the possible shifted edge users. Based on the estimated load increment, in the traffic shifting stage, the hot-spot cell tries to shift traffic and to keep each partner's load no exceed the load reported as available by the partner [LSJB10]. This scheme can reduce handover failure rate and improve user's satisfaction after shifting.

2.4.3.4 BS state analysis based MLB

As investigated in [HZZYW10] [JBTMK10], the handover procedure introduced by MLB consumes system signalling load and may impact networks performance in terms of handover failure and ping-pong handover.

In order to reduce the number of handovers introduced by frequent traffic shifting, Heng Zhang *et al.* designed a modified MLB in [ZQMZ10a]. The major contribution of [ZQMZ10a] is the novel BS state analysis and optimisation mechanism to reduce the number of handovers and to shift traffic effectively. This scheme [ZQMZ10a] includes four phases:

- Monitoring: To obtain the load information of local BS.
- Analysing: To analyse the state of BS. Each BS can be set as one of three states: "*high load*", "*normal load*", "*balancing*". "*High load*" state denotes a heavily loaded BS. "*Normal load*" state denotes that the BS is lightly loaded, and it's not receiving traffic (no response load balancing requirement [ZQMZ10a]) from a "*high load*" BS. "*Balancing*" state denotes that the BS is lightly loaded, while it is receiving traffic from a "*high load*" BS (response load balancing requirement [ZQMZ10a]).
- Optimisation: After obtaining the load information of neighbouring BSs via X2 interface, the "*high load*" cell selects "*normal load*" BSs as partners. Then the partners move into the "*balancing*" state. (To our knowledge, the above novel BS state analysis and optimisation mechanism guarantees that a lightly load BS can only receive the traffic from one heavily loaded BS at a time, thus reducing the number of handovers and improving the traffic shifting efficiency.)
- Implementation: The "*high load*" BS and its partner BSs ("*balancing*" BSs) adjust handover parameters and shift users, in order to balance load among cells.

2.4.4 Problem Formulation in Conventional MLB

The virtual partner problem and the aggravating load problem are discussed in this section, since they occur in conventional MLB schemes.

2.4.4.1 Virtual Partner Problem

In MLB, one of the most basic and important actions taken by the heavily loaded cell is to select suitable neighbouring cells as partner cells. In conventional MLB, the neighbouring cell's load is widely used as the criterion for finding partner cells. However, neighbouring cells with similar load may have different capabilities of serving the shifted users. Load based partner selection cannot effectively select partner and may lead to the virtual partner problem as shown in Figure 2.13.

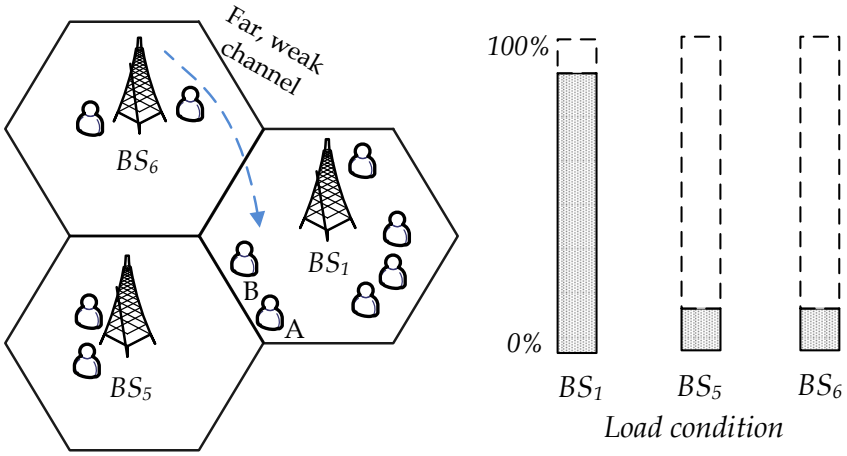


Figure 2.13 Virtual partner problem

In this simplified network, assuming each user requires the same amount of traffic; BS_1 is heavily loaded and intends to shift some traffic out. Applying the load criterion, both BS_5 and BS_6 appear to be possible partners with the same priorities as they have the same load. However, BS_6 cannot effectively serve $User_A$ and $User_B$ because they are far from BS_6. In this thesis, virtual partner is defined as a lightly loaded neighbouring cell, which is far from the heavily loaded cell's edge users. 'Virtual' means that this lightly loaded cell cannot effectively shift the heavily loaded cell's traffic. Hence, BS_6 is a virtual partner in Figure 2.13.

In order to effectively select partners and to address the virtual partner problem, a user-vote assisted clustering algorithm, which considers users' channel condition received from neighbouring cells, is proposed in Section 4.4.

2.4.4.2 Public Partner and Aggravating Load Problem

When multiple hot-spot BSs shift traffic to one partner, this partner becomes a *public partner*. Without the coordination of hot-spot BSs, their traffic may result in the public partner being heavily loaded. As shown in Figure 2.14, BS_5 is the public partner of both hot-spot BS_1 and BS_9. The amount of shifting traffic from each BS is moderate, while the total traffic from two BSs can result in heavily loaded BS_5.

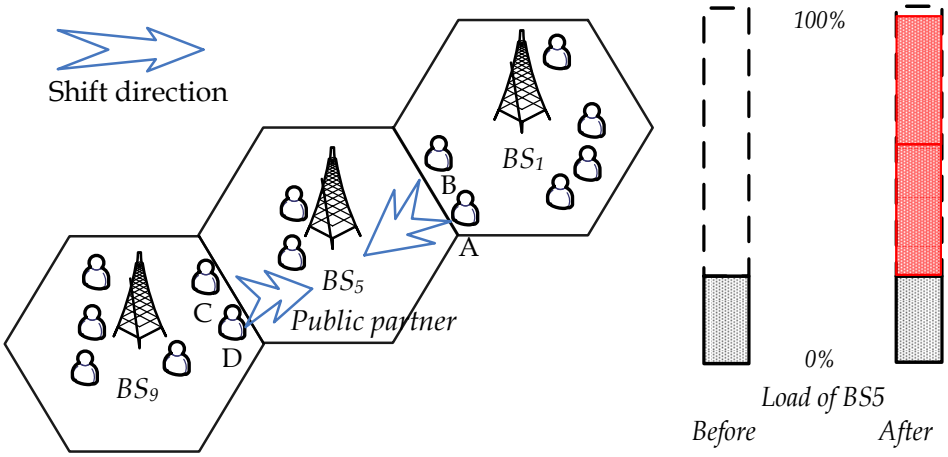


Figure 2.14 Aggravating load problem of public partner

In this thesis, the aggravating load problem means that the target node, which can be a public partner cell or a relay station (RS), becomes heavily loaded after traffic shifting. Therefore, the phenomenon of heavily loaded public partner is called the *aggravating load problem of public partner*. To our knowledge, a major reason of this problem is that a hot-spot cell cannot control other hot-spot cell's shifting traffic to the public partner, under distributed control LTE/LTE-Advanced networks.

The aggravating load problem impacts the public partner's performance. The public partner may reject the access requests/handover requests because it cannot provide a sufficient data rate to new call users/handover users; the public partner may be unable to provide existing users with the required QoS.

The MLB schemes in [NA07] [KAPTK10] [LLZL10] do not analyse the coordination of multiple hot-spot cells' traffic shifting towards one public partner. In [LSJB10], a hot-spot cell estimates the partner's load increment resulted from the possible shifted edge users, and then it tries to shift proper users to keep partner's load under the heavily loaded threshold. Hence, the scheme [LSJB10] can mitigate the *non-public partner* (receiving traffic from one hot-spot

cell) being heavily loaded, while it cannot effectively mitigate the heavily loaded *public partner*. This is because that in distributed control LTE/LTE-Advanced networks, the hot-spot cell cannot control the shifting traffic from other hot-spot cells to their public partners. In [ZQMZ10a], a lightly loaded cell can receive traffic from *only one hot-spot cell* at a time. This mechanism avoids the appearance of a heavily loaded public partner, while other hot-spot cells may lose the traffic shifting opportunity even though this lightly loaded cell has sufficient idle spectrum to assist other cells.

In order to coordinate multiple hot-spot cells' shifting traffic and to address the aggravating load problem, a relative load response model is proposed in Section 4.5.1.

2.5 Load Balancing in Multiple-Hop OFDMA Networks

In single-hop OFDMA networks, cell edge users may receive high inter-cell interference and low signal power. To address this issue, relay becomes a promising technology in future cellular networks, because relay can extend the cell coverage and enhance users' performance in cell edge area [Xiao10].

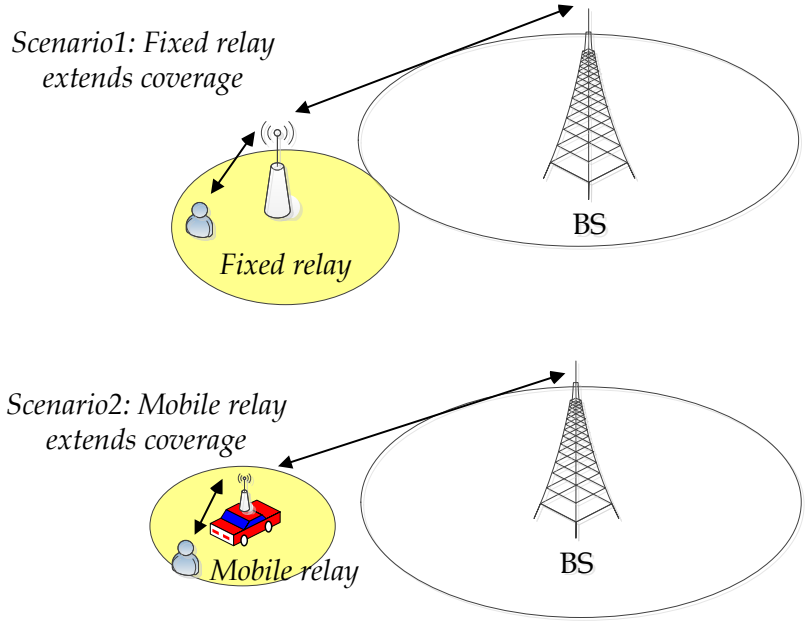


Figure 2.15 Two scenarios of relay extending cell coverage

As shown in Figure 2.15, there are two categories of relay: mobile relay and fixed rely. Compared with mobile relay, a fixed relay [GZLLZ07] [IEEE802web07] can achieve better cell edge coverage and a higher data rate. Therefore, 3GPP considers fixed relay as a

potential technology which will be adopted in LTE-Advanced networks [3GPP10e].

Although mobile relays may experience unstable channel condition, they have the advantage of flexibility and low configuration costs. Therefore, under emergency or disaster, mobile relay is needed to extend cell coverage flexibly.

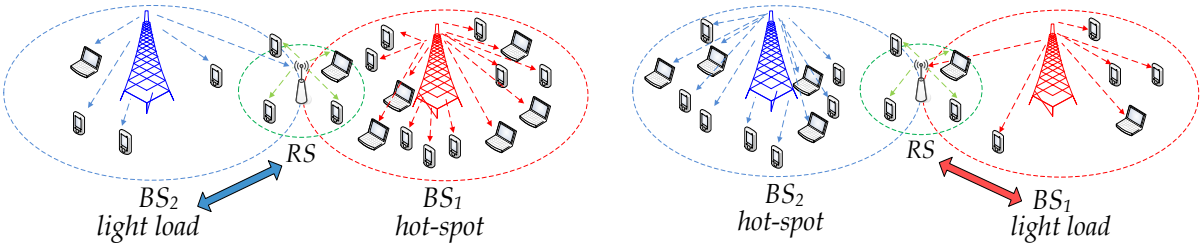
2.5.1 Load Balancing in Fixed Relay Cellular Networks

The appearance of fixed relay provides a new perspective to deal with the uneven load between neighbouring cells, e.g., shifting cell's load via relay stations (RSs) of neighbouring cells. This section presents two typical load balancing schemes in fixed relay cellular networks, including dynamic connection based load balancing scheme [JBW08] and traffic load balancing scheme [WTJLHL10].

1) Dynamic Connection based Load Balancing in Fixed Relay Cellular Networks

Peng Jiang *et al.* [JBW08] designed a dynamic connection based load balancing scheme in fixed relay cellular networks. Its basic idea is that RS has dynamic connection with neighbouring BSs, according to neighbouring cells' load. As exemplified in Figure 2.16(a), when $Cell_1$ is a hot-spot, an RS connecting to BS_1 will reconnect to lightly loaded neighbouring BS_2, and then the users served by this RS will be shifted to $Cell_2$. In Figure 2.16(b), when $Cell_2$ becomes a hot-spot, RS will switch back to the lightly loaded BS_1. Therefore, the dynamic BS-RS connection can release the hot-spot cell's spectrum resources and reduce the hot-spot cell's load.

The limitation of this scheme is that the dynamic BS-RS connection might result in a coverage gap, since BS needs time to configure and assign spectrum to new connected RS.



(a) Scenario1 (b) Scenario2

Figure 2.16 Illustration of dynamic BS-RS connection in [JBW08]

2) Traffic Load Balancing in Fixed Relay Cellular Networks

Due to the limitation of [JBW08], Xijun Wang *et al.* [WTJLHL10] designed a traffic load balancing scheme in fixed relay cellular networks. Its basic idea is that a hot-spot cell selects the neighbouring cell with the lowest load in sequence, until the average load of the hot-spot cell and selected partners is lower than the presetting threshold [WTJLHL10]. The hot-spot cell calculates the amount of shifting traffic, in order to make the traffic evenly distributed among the hot-spot cell and partners. According to the amount of the required shifting traffic and user's spectrum efficiency, the hot-spot cell offloads users to target RS in each partner. The traffic shifting of [WTJLHL10] is illustrated in Figure 2.17. Figure 2.17 assumes that $Cell_1$ selects $Cell_4$, $Cell_5$ and $Cell_7$ as partners. Since most shifted users are in $Cell_1$ edge, they will be served by target RSs of $Cell_4$, $Cell_5$ and $Cell_7$.

This scheme [WTJLHL10] can make use of RS to shift hot-spot cell's traffic effectively. However, this scheme does not consider the coordination of multiple hot-spot cells when they shift traffic to one public RS simultaneously, which might result in the RS aggravating load problem. The RS aggravating load problem is discussed in Section 2.5.3.

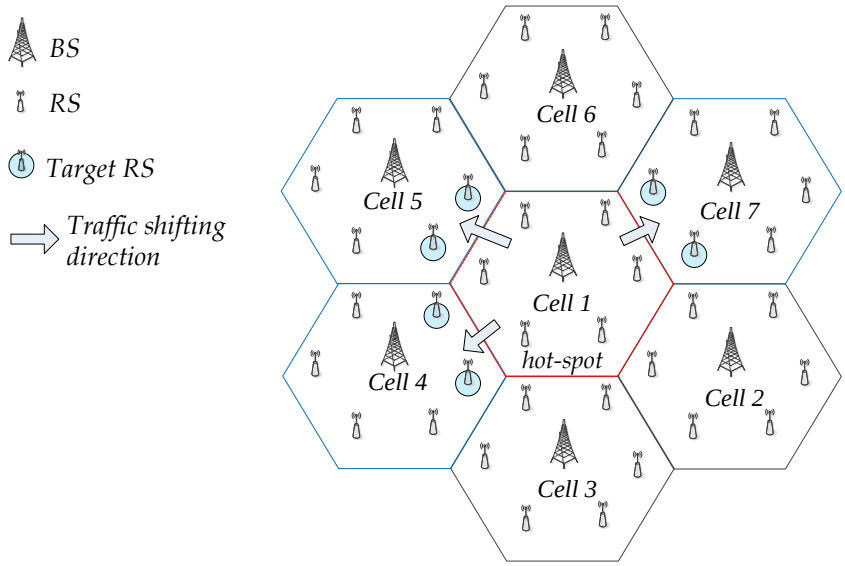


Figure 2.17 Illustration of traffic shifting in [WTJLHL10]

2.5.2 Load Balancing in Mobile Relay Cellular Networks

Mobile relay can be deployed in cellular networks flexibly. Therefore, mobile relay can extend lightly loaded cell's coverage to reduce hot-spot cell's load.

Hongyi Wu *et al*. [WDQYT05] investigated load balancing and proposed the integrated cellular and Ad-hoc relay (*iCAR*) scheme. In the *iCAR* scheme, the ad-hoc relaying stations (ARS) are placed at strategic locations and employed to relay the traffic from a hot-spot cell to lightly loaded neighbouring cells [WDQYT05]. As exemplified in Figure 2.18. BS_1 is a hot-spot and BS_2 is lightly loaded. If a user requests access to hot-spot BS_1, it will be blocked by BS_1. Using the *iCAR* scheme, this user can be shifted to lightly loaded BS_2, via multi-hop communication through ARS_1 and ARS_2.

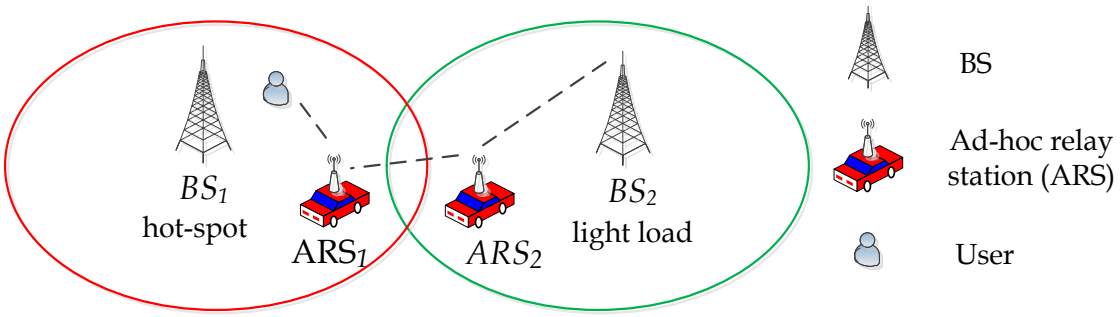


Figure 2.18 Example of load balancing via iCAR scheme in [WDQYT05]

2.5.3 Challenges of Load Balancing in Multi-Hop Networks

The deployment of fixed/mobile relay increases the complexity of load balancing in cellular networks. In fixed relay cellular networks, load balancing may result in the RS aggravating load problem. In mobile relay cellular networks, mobile relay selection directly impacts user performance.

- RS aggravating load problem: Most of users shifted from a hot-spot cell will be served by adjacent RSs in partner cells, since RSs are always located at the cell edge to provide shifted users with good link quality. However, each RS is pre-allocated only a small part of spectrum resources in the partner cell. Therefore, the shifted users may result in the RS aggravating load problem, especially under public RS scenario, in which multiple hot-spot cells shift traffic to a RS simultaneously and this RS becomes public RS.

  As exemplified in Figure 2.19, $Cell_1$ and $Cell_8$ are heavily loaded and try to shift some traffic to the lightly loaded $Cell_2$. Due to the random user distribution, most edge users of $Cell_1$ and $Cell_8$ will be shifted to RS_1 of $Cell_2$. After receiving traffic, RS_1 becomes heavily loaded. In this thesis, the phenomenon of heavily loaded RS is denoted as the *RS aggravating load problem*. This problem impacts networks performance because

heavily loaded RS_1 may result in the handover failure. In order to address the RS aggravating load problem, a relay-level user shifting algorithm is designed in Section 5.2.6.

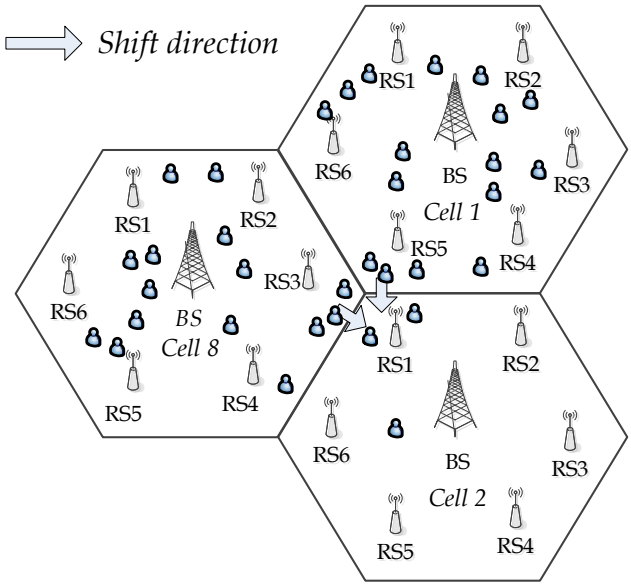


Figure 2.19 RS aggravating load problem

- Mobile relay selection: Relay selection impacts user performance due to different relay channel conditions. In Section 5.3, user relaying assisted traffic shifting (URTS) scheme is designed. The URTS scheme can select an appropriate mobile relay to improve the shifted user's link quality under low cost of mobile relay's energy consumption.

2.6 Summary

This chapter introduces the background of load balancing, including its application scenario, process, and the categories of load balancing schemes. Then the evolution of load balancing schemes from 2G to the state-of-the-art MLB schemes proposed for the latest LTE/LTE-Advanced networks is discussed. Finally, load balancing schemes and challenges in multiple-hop OFDMA networks are discussed.

Chapter 3 System Model and Simulation Platform

This chapter describes the system model and the simulator platform. The system model is the foundation for designing novel load balancing schemes in later chapters. The simulation platform is the foundation for the performance analysis to evaluate the proposed schemes.

3.1 System Model

This thesis researches load balancing in three types of system models. Each system model adopts OFDMA as the multiple access technology.

3.1.1 Single-Hop Cellular Networks

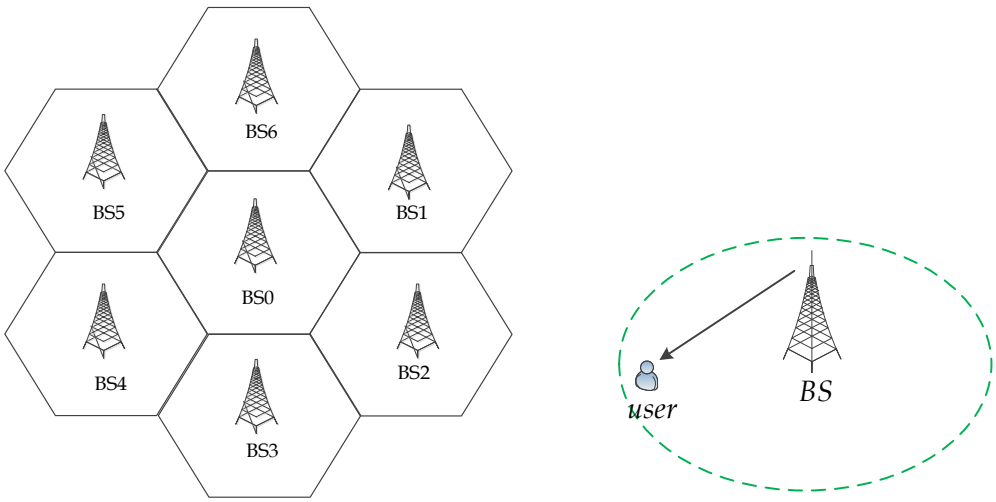


Figure 3.1 Single-hop cellular networks

The first system model is single-hop cellular networks [3GPP08a] [3GPP10c] [RY10]. As shown in Figure 3.1, BS transmits signal to user directly. This system model employs full

frequency reuse (FRF=1) technology with the carrier frequency of 2GHz. It also employs the LTE macro-cell propagation model [3GPP11c].

3.1.2 User Relay Cellular Networks

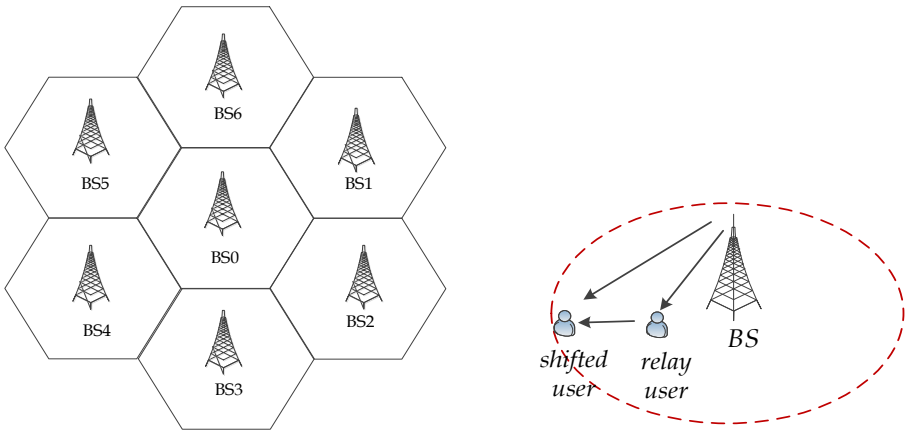


Figure 3.2 User relay cellular networks

The second system model is user relay cellular networks. In this system model, there are both active users and non-active users. When a user is in idle mode [3GPP10c] [3GPP11d], it is called the non-active user in this thesis. As depicted in Figure 3.2, besides the BS direct link, the non-active user is employed as the relay to forward signal for the active shifted user. The relay transmission mode is introduced in Section 5.3.2. The basic cellular parameters in this system model are the same as those in the first system model, such as frequency planning, carrier frequency, bandwidth, BS location, and BS transmit power.

3.1.3 Fixed Relay Cellular Networks

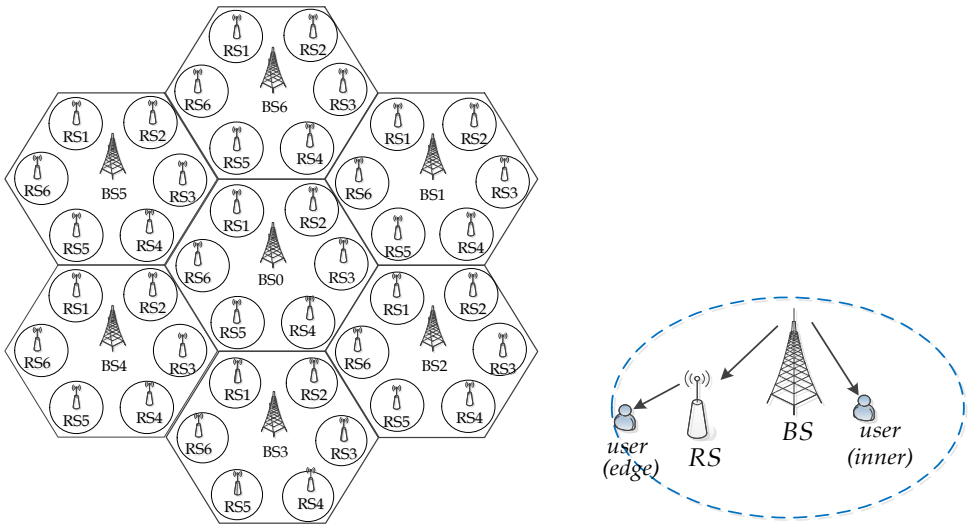


Figure 3.3 Fixed relay cellular networks

The third system model is fixed relay cellular networks, as shown in Figure 3.3. The user located in cell inner area is served by BS directly. Meanwhile, the user located in cell edge area is served by relay station (RS) via two-hop transmission. RS works in decode-and-forward (DF) mode [WJL09]. The frequency planning refers to the modified soft frequency reuse (MSFR) technology [GZLLZ07], which can achieve full frequency reuse among cells. Correspondingly, one OFDMA system level simulation platform is built. This simulation platform includes above three system models.

3.2 Overall Design of Simulation Platform

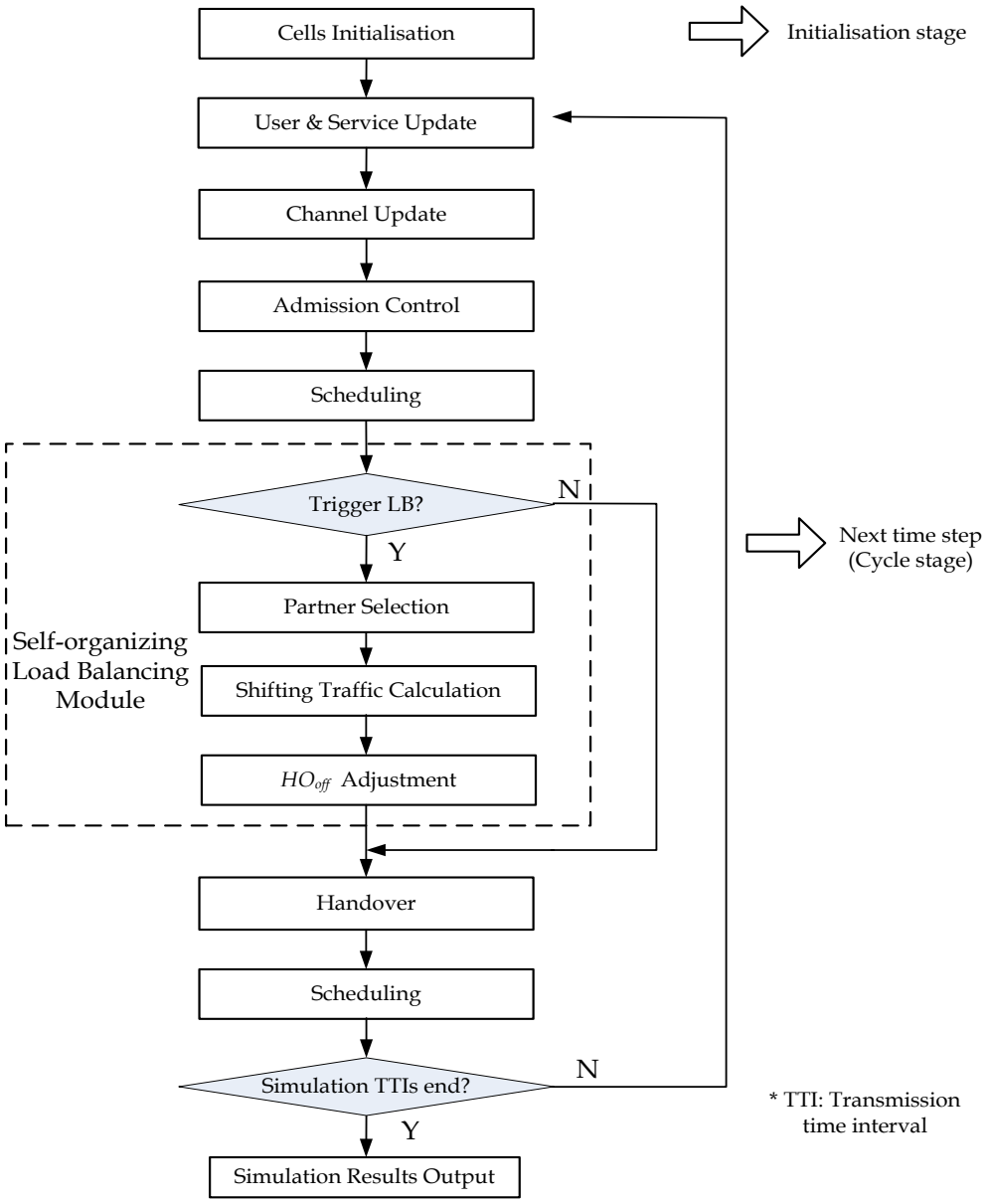


Figure 3.4 Flowchart of simulation platform

Figure 3.4 shows the flowchart of the system level simulation platform. This simulation platform uses time-stepping. The function of each module is as follows:

Cells initialisation

This module initialises the system parameters, including the network topology, the position of BS and RS, the antenna configuration, and the frequency planning.

User and service update

User location: This module generates both active users and non-active users, and then initialises their physical locations. Since this simulation platform uses time-stepping, users' physical locations are updated at every time step.

Service: In this simulation platform, each active user adopts the GBR service [ANHV10]. GBR service can simulates the scenario that a cell's traffic load increase with the increase of the number of active users. In addition, the non-active users have no service requirements in this simulation platform.

Channel update

This module updates the path-loss according to users' physical locations. This module also updates the shadow fading in cellular networks.

Admission control

Admission control is a decision module in which the output is a yes/no decision about whether to admit user's access request or not. In this simulation platform, a new call user selects a cell from which the user receives the strongest reference signal received power (RSRP). When the number of subcarriers is enough to meet the user's service requirement, the user can access [KAK10]. Otherwise, the new call user will be blocked.

Handover module

Handover is a crucial module to keep the users' mobility and the seamless communication in cellular networks. RSRP based hard handover [ACMRP07] [3GPP12] is used in this simulation platform. The handover condition is presented in Equation (2.2) of Section 2.4.2.

Scheduling algorithm

After the handover module and the admission control module, the cell employs scheduling algorithm to allocate subcarriers to serving users.

Self-organising load balancing

This module implements the proposed self-organising cluster-based cooperative load balancing scheme. The proposed scheme is presented with details in Chapter 4. In Section 4.6, its performance is evaluated in single-hop cellular networks. Then, the proposed scheme is modified to be feasible in both fixed relay cellular networks and user relay cellular networks, which will be introduced in Section 5.2 and Section 5.3, respectively.

3.3 Cells Initialisation

3.3.1 Cells Initialisation in Single-Hop and User Relay Cellular Networks

The cells initialisation module of single-hop cellular networks is the same as that of user relay cellular networks. The flowchart of this module is shown in Figure 3.5.

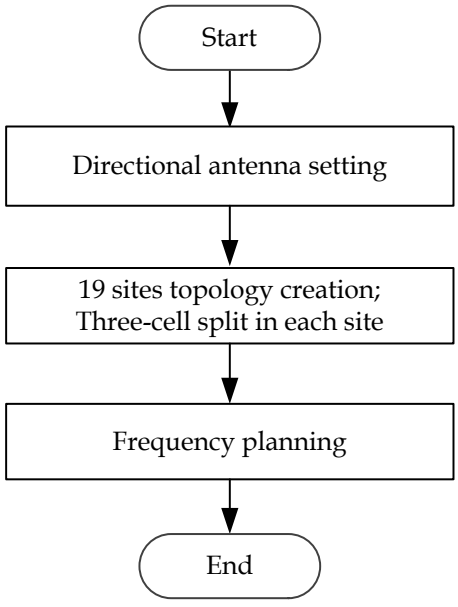


Figure 3.5 Flowchart of cell initialisation in single-hop and user relay cellular networks

1) *Directional Antenna*

In the cell initialisation module, the directional antenna is employed to split each site into

three cells. According to [3GPP08c], the three-sector antenna pattern is formulated as

$$A(\theta) = -\min\{12(\frac{\theta}{\theta_{3dB}})^2, A_m\} \tag{3.1}$$

where $\theta_{3dB} = 70^\circ$ (degree), $A_m = 20dB$.

2) *19 Sites Topology and Three-Cell Division*

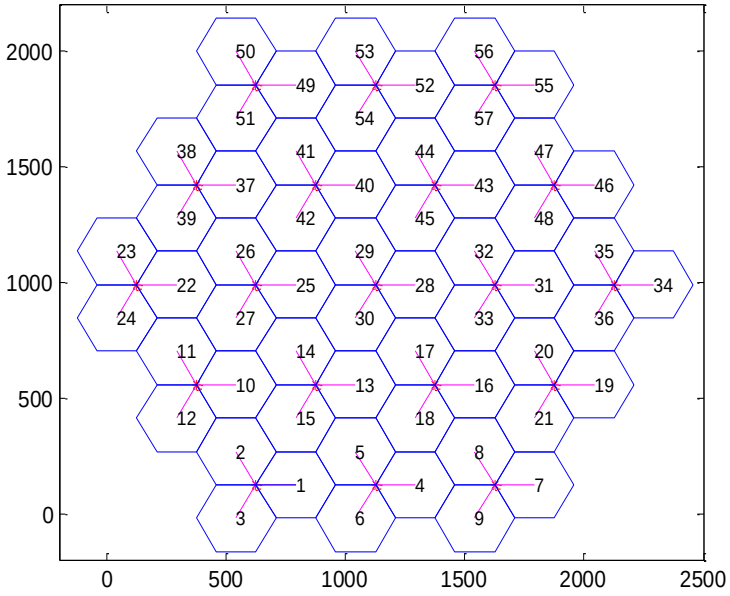


Figure 3.6 Simulated cell layouts of single-hop cellular networks (unit: *meter*)

The three-sector antenna based multi-cell scenario is shown in Figure 3.6. There are 19 sites with 57 cells, and the transmitter of each cell is located in the interchanging point of each site. The inter-site distance is 500*meters*. The sites topology and three-cell division meet the requirement of system design in 3GPP (e.g., Table A.2.1.1-1 in [3GPP10d]). This is also adopted by related works, such as [KAK10] [KAPTK10] [LSJB10].

3) *Frequency Planning*

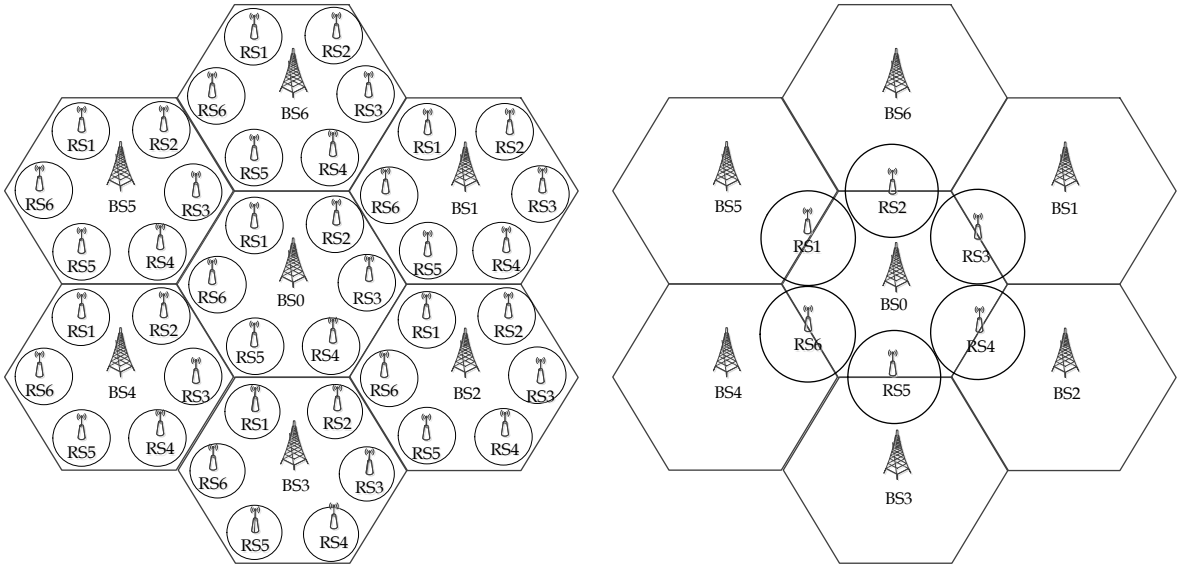
Due to high rate requirements of future wireless networks, 3GPP considers that each cell could use the whole spectrum resources, which is denoted as the full frequency reuse technology [3GPP10c] [RY10]. Besides, 3GPP LTE supports different types of spectrum bandwidth, including 1.4MHz, 3MHz, 5MHz, 10MHz, 15MHz and 20MHz. In the single-hop cellular networks simulator, the spectrum bandwidth is set as 5MHz, which is used in

related works, such as [KAK10] [KAPTK10] [YLCW12]. This can help the performance comparison between the proposed CCLB scheme and related MLB works.

3.3.2 Cells Initialisation in Fixed Relay Cellular Networks

In fixed relay cellular networks, the cell initialisation module initialises the position of BS and RS, and pre-allocates subcarriers to each BS and RS.

1) *Cells Topology and Fixed Relay Location*



(a) RSs located at 2/3 of cell radius (b) RSs located at cell boundary

Figure 3.7 Two types of relay location

As shown in Figure 3.7, there are two cell structures according to the location of fixed RSs. Sunghyun Cho *et al.* [CJC09] proved that the cell structure of RSs located at 2/3 cell radius (Figure 3.7(a)) can provide higher overall system rate than the structure of RSs located at cell boundary (Figure 3.7(b)). Therefore, the cell structure in Figure 3.7(a) is adopted in this simulation platform. Figure 3.8 shows the simulated cell layouts of fixed relay cellular networks scenario. There are 19 cells with the inter-BS distance of 1500m [ZFLLW11], and hence the cell radius is 877m. In each cell, 6 RSs are located at 2/3 of the cell radius. The total transmit power of BS is 46dBm, and that of RS is 37dBm [ZFLLW11].

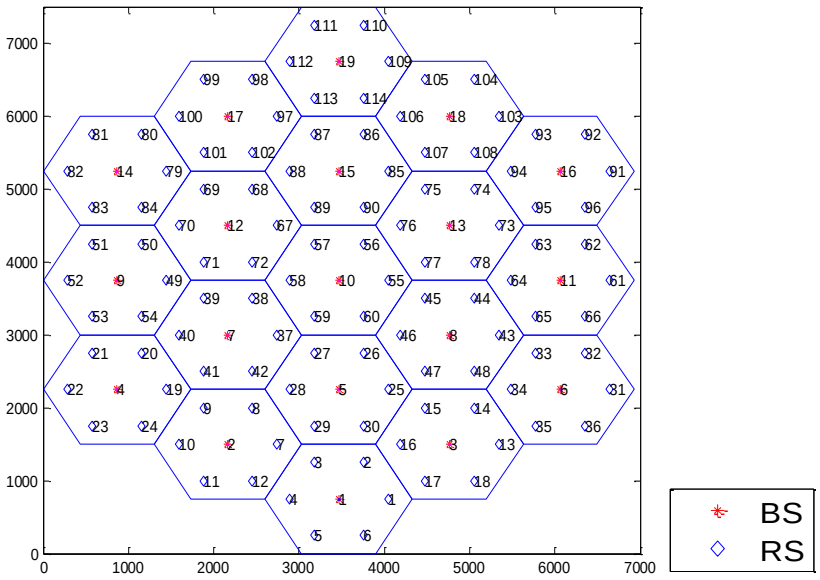


Figure 3.8 Simulated cell layouts of fixed relay cellular networks

2) *Frequency Planning and Transmission Mode*

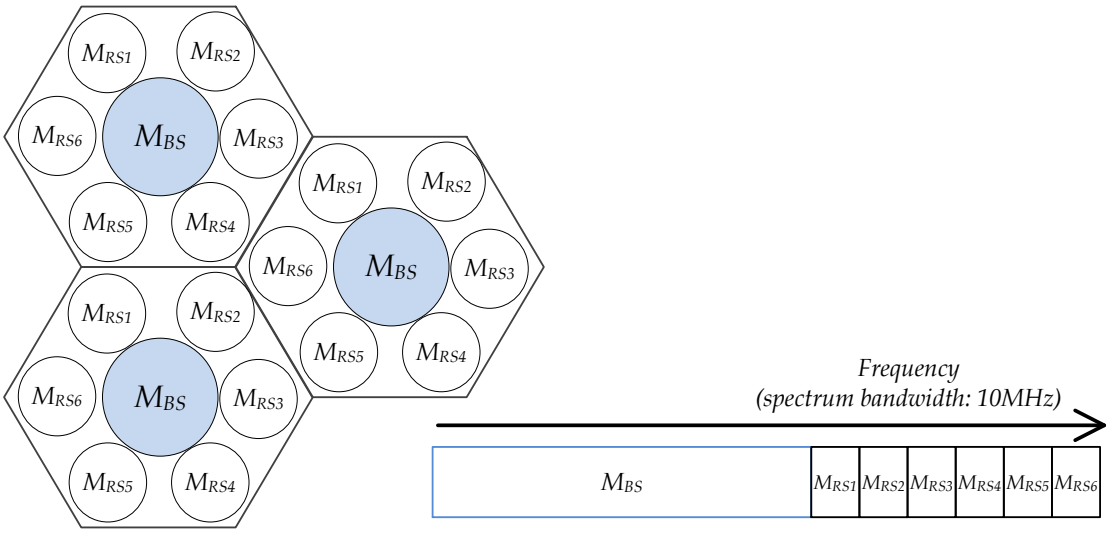


Figure 3.9 Illustration of MSFR technology [GZLLZ07]

This simulator employs the modified soft frequency reuse (MSFR) technology, which is designed in [GZLLZ07]. As illustrated in Figure 3.9, in the simulator of fixed relay cellular networks, the spectrum bandwidth is *10MHz*. The *10MHz* spectrum is reused in all cells. According to [GZLLZ07], in each cell, the total subcarriers are divided into two orthogonal sets. Their numbers of subcarriers are M_{BS} and M_{RS}, respectively. BS can use M_{BS} subcarriers. Furthermore, M_{RS} subcarriers are divided into six equal groups of subcarriers, as M_{RS1},M_{RS2}….M_{RS6}. Each RS can use a corresponding group of subcarriers, e.g., RS_4 can use M_{RS4} subcarriers.

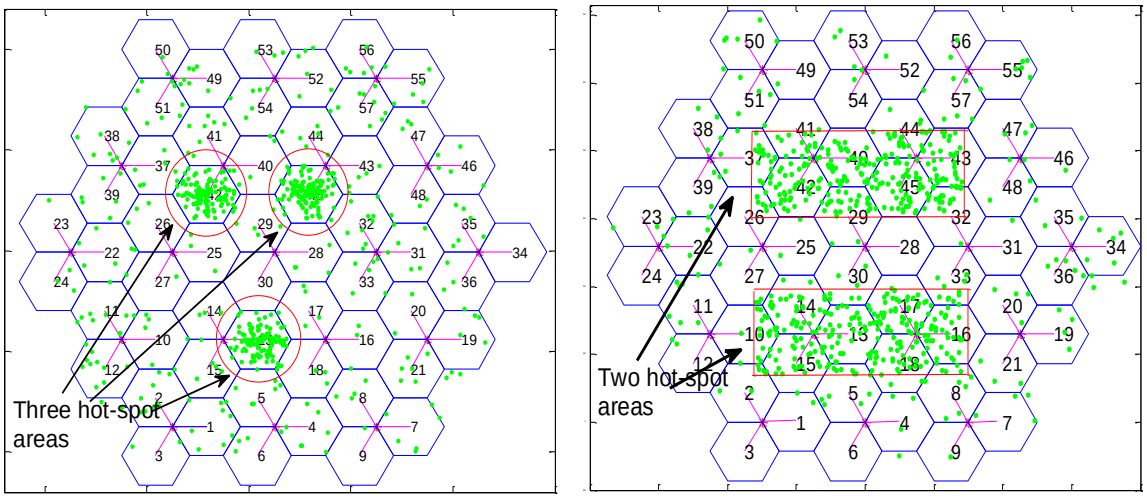
BS transmits data to the inner user with M_{BS} subcarriers. For the edge user, the two-hop transmission mode refers to [GZLLZ07]: the downlink transmission is split into two consecutive stages. In the first stage, BS transmits signal to RS with M_{RS} subcarriers. In the second stage, RS decodes the signal, and forwards the re-encoded signal to the edge user with M_{RS} subcarriers.

3.4 User Distribution

Since load balancing aims at dealing with the uneven load distribution, this simulation platform generates unevenly user distribution scenarios for load balancing performance evaluation.

3.4.1 User Distribution in Single-Hop Cellular Networks

The simulator can generate different types of hot-spot areas. This simulator can set the number and the shape of hot-spot areas. For example, three circle hot-spot areas and two rectangle hot-spot areas are generated in Figure 3.10(a) and Figure 3.10(b), respectively.



(a) Six circle hot-spot areas (b) Two rectangle hot-spot areas

Figure 3.10 Uneven users distribution in single-hop cellular networks

3.4.2 User Distribution in User Relay Cellular Networks

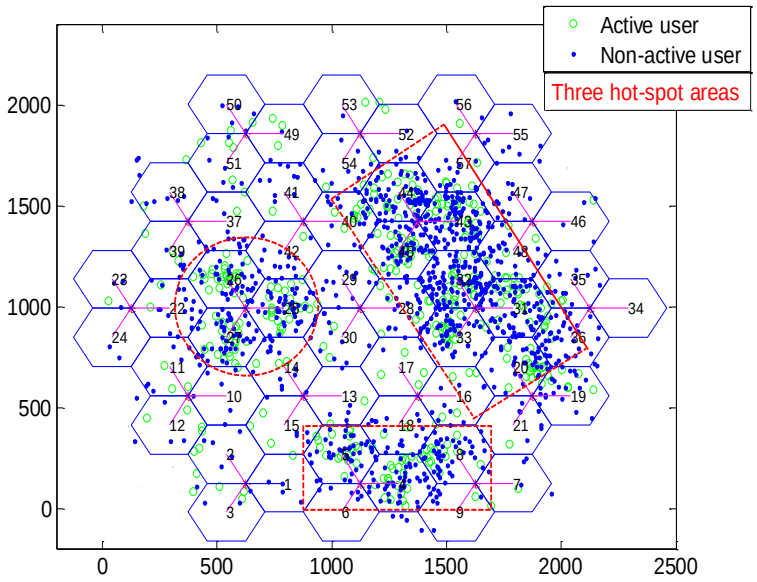


Figure 3.11 Uneven users distribution in user relay cellular networks

The simulator of user relay cellular networks can generate active users and non-active users [3GPP10c]. The two types of users have a similar distribution. Figure 3.11 shows that most active users (green circle) and non-active users (blue dot) are in three hot-spot areas.

3.4.3 User Distribution in Fixed Relay Cellular Networks

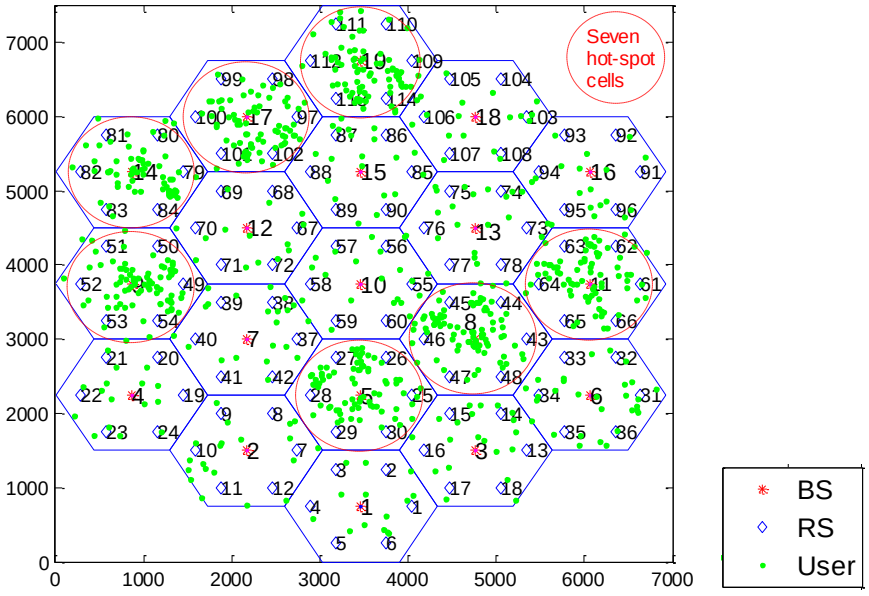
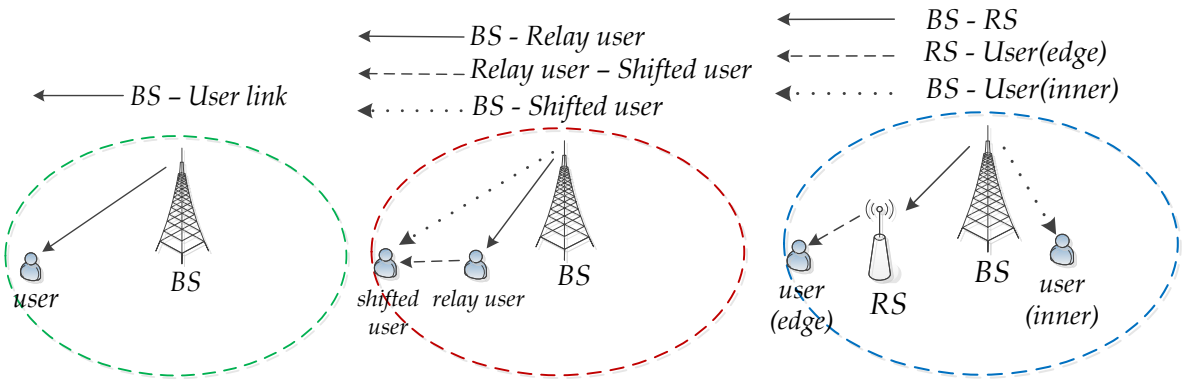


Figure 3.12 Uneven users distribution in fixed relay cellular networks

The simulator of fixed relay cellular networks can also generate user distribution unevenly. Figure 3.12 exemplifies seven hot-spot cells in networks.

3.5 Channel Model

In a wireless communication system, the channel state plays an important role in determining the system performance, and it affects the quality of communication. In order to rapidly and accurately simulate channel state, this module updates the path-loss according to each user's physical position. Meanwhile, the shadow fading is also updated in this module. The downlinks of three system models are shown in Figure 3.13.



(a)Single-hop cellular networks (b)User relay cellular networks (c)Fixed relay cellular networks

Figure 3.13 Downlinks in three system models

3.5.1 Channel Model in Single-Hop Cellular Networks

In single-hop cellular networks, the large scale path-loss model refers to LTE macro-cell propagation model [3GPP11c] [IEEE802web11], which is applicable for urban areas and can be express as

$$L_{BS\text{-}User}(\text{dB}) = 40 \times (1 - 4 \times h_b) \times \log_{10}(d_{BS-User}) - 18 \times \log_{10}(10^3 \times h_b) + 21 \times \log_{10}(f_c) + 80 \quad (3.2)$$

where $d_{BS-User}$ (unit: *kilometer*) is the distance between BS and user, f_c is the frequency and this simulator sets $f_c = 2000MHz$, h_b is the height of the BS and this simulator sets $h_b = 0.015km$. Therefore, Equation (3.2) can be expressed as (3.3). This path-loss model (3.3) is widely used in LTE/LTE-Advanced [3GPP10d] [HGV10] [3GPP11c] [RH12].

$$L_{BS\text{-}User}(\text{dB}) = 37.6\log_{10}(d_{BS-User}) + 128.1 \qquad d_{BS-User}(km) \quad (3.3)$$

BS is located on the rooftop and user is below the rooftop. BS-User link may suffer shadow fading due to obstacles, such as building. The level of shadow fading is usually simulated by using a log-normal distributed random variable, and the typical log-normal shadow fading model [Xiao10] is used here. The standard deviation value is set as 8dB under non-line-of-sight (NLOS) transmission [3GPP10d].

3.5.2 Channel Model in User Relay Cellular Networks

As shown in Figure 3.13(b), the user relay cellular networks have three types of links. Since the user relay cellular networks simulator is based on the single-hop networks simulator, the path-loss models of the BS-Relay user link and the BS-Shifted user link are shown in (3.4a) and (3.4b), respectively.

$$L_{BS\text{-}Relay\ user}(\text{dB}) = 37.6\log_{10}(d_{BS\text{-}Relay\ user})+128.1 \qquad d_{BS\text{-}Relay\ user}(km) \tag{3.4a}$$

$$L_{BS\text{-}Shifted\ user}(\text{dB}) = 37.6\log_{10}(d_{BS\text{-}Shifted\ user})+128.1 \qquad d_{BS\text{-}Shifted\ user}(km) \tag{3.4b}$$

$$L_{Relay\ user\text{-}Shifted\ user}(\text{dB}) = b_0 + 20\times\log_{10}(f_c) + 10\times n\times\log_{10}(d_{Relay\ user\text{-}Shifted\ user}) \tag{3.4c}$$

where the unit of f_c is *MHz*, the unit of $d_{Relay\ user\text{-}Shifted\ user}$ is *meter* in (3.4c)

In this simulation platform, the path-loss model of Relay user-Shifted user link is shown in (3.4c) [Damosso98] [WTN04]. This path-loss model suits inter-user communication scenario with frequency f_c at *2000MHz*. From [WTN04], the values of b_0 and n depend on specific application scenario and antenna setting, e.g., b_0 can be *-27.6* and n can be *2.0* [WTN04].

In user relay cellular networks simulator, three types of downlinks all consider the log-normal shadow fading with 8dB standard deviation.

3.5.3 Channel Model in Fixed Relay Cellular Networks

1) *BS-RS Link*

Figure 3.14 shows the transmission scenario between BS and RS (Figure 2 in

[IEEE802web07]). Since RS is located on the rooftop, it is line-of-sight (LOS) between BS and RS [IEEE802web07] [3GPP10d]. Therefore, in this simulator, BS-RS link considers the LOS path-loss model without shadow fading. According to [IEEE802web11], this simulator sets BS's height $h_b = 0.015km$, RS's height $h_r = 0.012km$, and frequency $f_c = 2000MHz$.

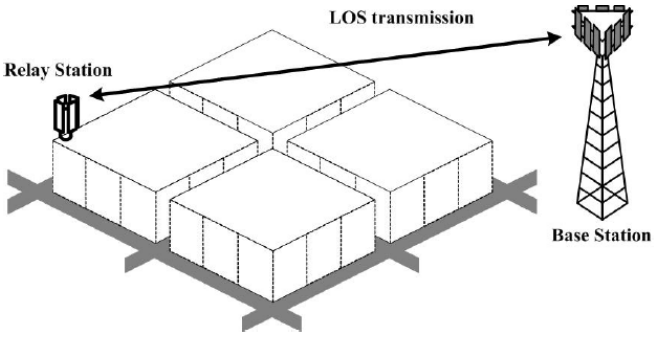


Figure 3.14 BS-RS LOS transmission [IEEE802web07]

The BS-RS LOS path-loss model is calculated as [3GPP10d]

$$L_{\text{BS-RS}}(\text{dB}) = 100.7 + 23.5 \times \log_{10}(d_{BS\text{-}RS}) \qquad d_{BS\text{-}RS}(km) \tag{3.5}$$

2) *RS-Edge user Link*

RS is located on the rooftop and user is below the rooftop. Hence, RS-Edge user link suffers NLOS transmission and shadow fading due to obstacles. The fixed relay cellular networks simulator employs the log-normal shadow fading with 8dB standard deviation. This simulator also adopts LTE macro-cell propagation model, as shown in (3.6a) [3GPP09] [3GPP11c]. Under $h_r = 0.012km$ and $f_c = 2000MHz$, the RS-Edge user path-loss model is as (3.6b).

$$L_{RS\text{-}Edge\,user}(\text{dB}) = 40 \times (1 - 4 \times h_r) \times \log_{10}(d_{RS\text{-}Edge\,user}) - 18 \times \log_{10}(10^3 \times h_r) + 21 \times \log_{10}(f_c) + 80 \tag{3.6a}$$

$$\Rightarrow \qquad L_{RS\text{-}Edge\,user}(\text{dB}) = 38.1 \times \log_{10}(d_{RS\text{-}Edge\,user}) + 129.9 \qquad d_{RS\text{-}Edge\,user}(km) \tag{3.6b}$$

3) BS-Inner user Link

For the cell inner user served by BS via single-hop link, the channel model is similar to that in Section 3.5.1: the path-loss model is shown in Equation (3.3); the shadow fading employs the log-normal shadow fading with 8dB standard deviation.

3.6 Scheduling

Scheduler allocates spectrum resources to users. Its aim is to effectively use spectrum resources and improve the networks performance. This section introduces the structure of physical resource block (PRB), which is the basic allocation unit in the scheduler of OFDMA cellular networks. Then the basic idea of the maximum carrier to interference ratio (Max-C/I) scheduler is described.

1) *Physical Resource Block*

According to 3GPP LTE type-1 frame structure [3GPP08a] [3GPP10c], each 10ms radio frame is divided into 10 equally sized sub-frames, and the length of each sub-frame is the 1ms transmission time interval (TTI). Each sub-frame consists of 2 equally sized time slots. The sub-frame structure is shown in Figure 3.15 and a sub-frame is composed of 2 PRBs.

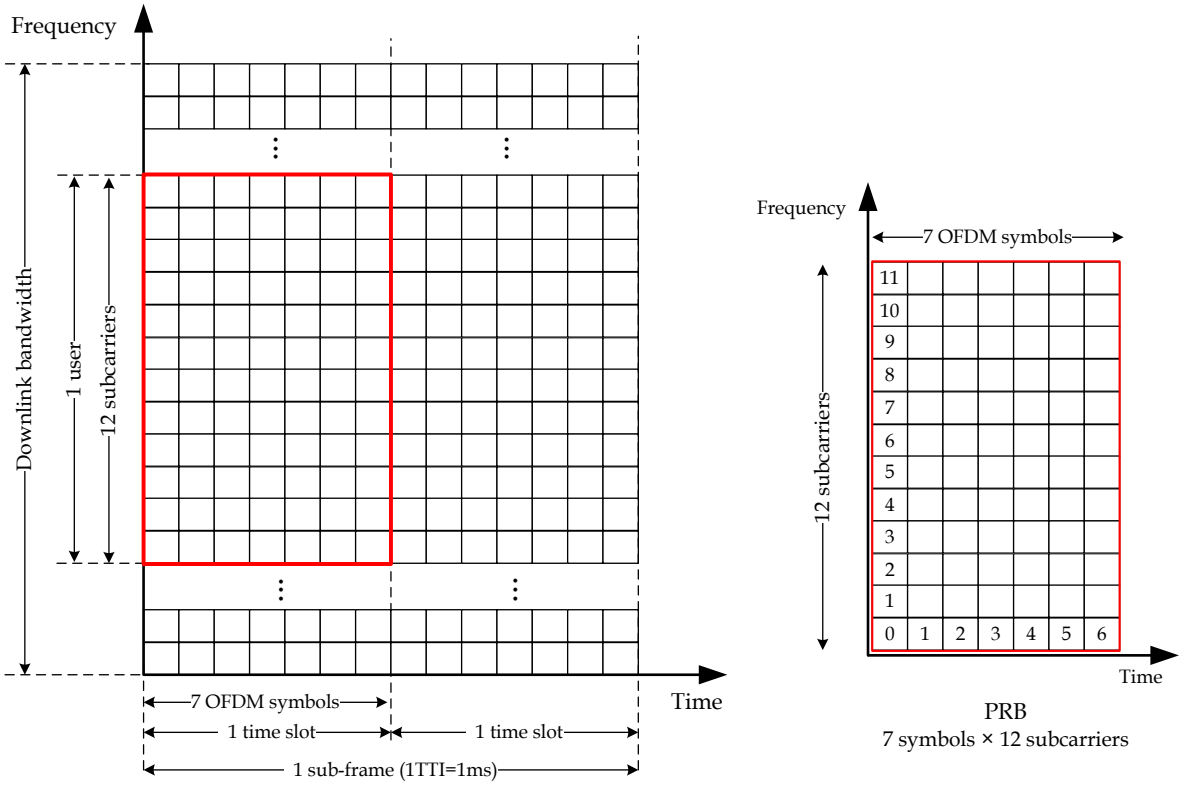


Figure 3.15 Sub-frame structure and PRB structure in [3GPP08a]

As shown in Figure 3.15, in the time domain, each PRB includes *7* OFDM symbols; in the frequency domain, each PRB contains *12* subcarriers (*15KHz* per subcarrier) [3GPP08a]. Therefore, the bandwidth of each PRB is *180KHz*. In the scheduling module of OFDMA

cellular networks, the PRB is the basic allocation unit. In each cell, the scheduling module can allocate PRBs to users at each scheduling period. The scheduling module is introduced as follows.

2) *Max-C/I Scheduler*

The channel conditions of different users' radio links vary independently. At each scheduling period, there always exists a radio link whose channel condition is the peak among all links. Thus, BS can allocate spectrum resources to this radio link for transmission, which will achieve a high data rate.

Scheduling the user with the instantaneously best radio link is Max-C/I scheduler. From the system rate perspective, Max-C/I scheduler can achieve the highest system rate among all types of schedulers [WZ05] [KSK08]. The Max-C/I scheduler can be expressed as scheduling *User l* given by

$$User\ l = \arg \max_{k \in \{1 \ldots K\}} R_k \tag{3.7}$$

where R_k is the instantaneous data rate for *User k* ($k \in \{1,2 \ldots K\}$). In this simulation platform, each cell allocates PRBs to users per scheduling period.

3.7 Overall Simulation Parameters

3.7.1 Parameters in Single-Hop Cellular Networks

The simulator of single-hop OFDMA cellular networks is designed based on 3GPP documents [3GPP08a] [3GPP08c] [3GPP10d] [3GPP11c] and related works, such as [ACMRP07] [HGV10] [KAK10] [KAPTK10] [LGK10] [LSJB10] [SOCRATES10] [RH12]. The simulation parameters are shown in Table 3.1.

Table 3.1 Simulation parameters of single-hop cellular networks

Parameter	Value
Number of Sites	19
Number of Cells per Site	3
Inter-Site Distance	500m
Transmission Time Interval (TTI)	1 ms
Number of Subcarriers	300
Number of PRBs	25 (12 subcarriers/PRB)
PRB Bandwidth	180KHz per PRB, 15KHz per OFDM subcarrier
Carrier Frequency	2.0GHz
Total Bandwidth	5MHz
Antenna Pattern (Directional Antenna)	$A(\theta) = -\min\{12(\frac{\theta}{\theta_{3dB}})^2, A_m\}$ $\theta_{3dB} = 70^\circ$ (degree), $A_m = 20dB$
Antenna Gain	14dBi
BS Total Transmit power	43dBm
Scheduler	Max-C/I
User Mobility	Speed: 5m/s. Direction: Random
Service	64Kbps GBR per active user
BS-to-User Path-Loss Model	$L_{BS\text{-}User}(\text{dB}) = 37.6\log_{10}(d_{BS-User})+128.1$
Fading Model	Lognormal shadow fading model Standard Deviation: 8dB

3.7.2 Parameters in User Relay Cellular Networks

Table 3.2 shows simulation parameters of user relay cellular networks. This simulator is based on single-hop cellular networks simulator, and the difference is that the user relay cellular networks simulator has non-active users related parameters and user relaying link, which are designed based on related works, such as [Damosso98] [WTN04] [CYOCY08] [JXJA08] [WJL09] [3GPP10c] [WTJLHL10].

Table 3.2 Simulation parameters of user relay cellular networks

Parameter	Value
Number of PRBs	25 (12 subcarriers/PRB)
Total Bandwidth	5Mhz
Carrier Frequency	2GHz
Inter-Site Distance	500m
BS Total Transmit Power	43dBm
Non-active User Transmit Power	21dBm
Relay mode	Amplify-and-forward
Path-Loss Model of Three Links	See Section 3.5.2
Fading Model	Lognormal shadow fading model Standard Deviation: 8dB
BS Antenna Pattern (Directional Antenna)	$A(\theta) = -\min\{12(\frac{\theta}{\theta_{3dB}})^2, A_m\}$ $\theta_{3dB} = 70^\circ$ (degree), $A_m = 20dB$
BS Antenna Gain	14dBi
Non-active User Antenna	Omni-directional antenna
Mobility of Active /Non-Active User	Speed: 3m/s. Direction: Random
Scheduler	Max-C/I
Service of Active /Non-Active User	64Kbps GBR per active user 0 PRB per non-active user

3.7.3 Parameters in Fixed Relay Cellular Networks

The simulator of fixed relay cellular networks is based on 3GPP documents [3GPP08a] [3GPP10c] [3GPP10d] [3GPP11c] and related works, such as [ACMRP07] [GZLLZ07] [IEEE802web07] [CJC09] [KAK10] [LGK10] [SOCRATES10] [ZFLLW11] [ZWLZB12]. The simulation parameters are shown in Table 3.3.

Table 3.3 Simulation parameters of fixed relay cellular networks

Parameter	Value
Number of Cell	19
Number of BS	19
Number of RS	114 (19*6)
Cell Radius	877m (inter-BS distance 1500m)
Distance between BS and RS	2/3 Cell Radius
Carrier Frequency	2 GHz
Total Bandwidth	10MHz
Number of PRBs	50 (12 subcarriers/PRB)
PRB Bandwidth	180KHz per PRB, 15KHz per subcarrier
Frequency Planning	Modified soft frequency reuse BS: 26PRBs; RSs: 24PRBs (each RS: 4PRBs)
BS Antenna	Three-sector directional antenna
RS Antenna	Omni-directional antenna
BS Transmit Power	46dBm
RS Transmit Power	37dBm
Relay mode	Decode-and-forward
Scheduler	Max-C/I
User Mobility	Speed: 5m/s. Direction: Random
Service	64Kbps GBR per active user
Path-Loss Model	See Section 3.5.3
Fading Model	Lognormal shadow fading model Standard Deviation: 8dB

3.8 Simulation Iteration

In this simulation platform, each module is designed according to 3GPP documents and related works. In addition, this simulation platform refers to the open source simulator 'LTE Downlink System Level Simulator' [TUWienLTEsimulator].

In single-hop cellular networks, the sites topology and three-cell division are validated in Section 3.3.1. In fixed relay cellular networks, the relay location is validated in Section 3.3.2. In the simulation platform, the user distribution is introduced in Section 3.4, which satisfies the requirement of uneven load. The channel model is introduced in Section 3.5, which is

designed according to 3GPP documents. The radio resource management functionalities, such as admission control module and handover module, are designed according to related works. After designing all modules, we track a user in this simulation platform, in this way, we validate all modules can work together and validate the parameters and outputs in each module follow the system design requirements.

Apart from each module's validation, the number of simulation iterations impacts the simulation results. This is because that the actual locations of users will change in each simulation loop. From the simulation point of view, the effect of randomness and the average networks performance can be achieved by using a large number of simulation iterations. However, a larger number of simulation iterations take long time to get results. In order to choose a reasonable number of simulation iterations, networks overall blocking probability is compared under different numbers of simulation iterations, as shown in Figure 3.16 and Figure 3.17.

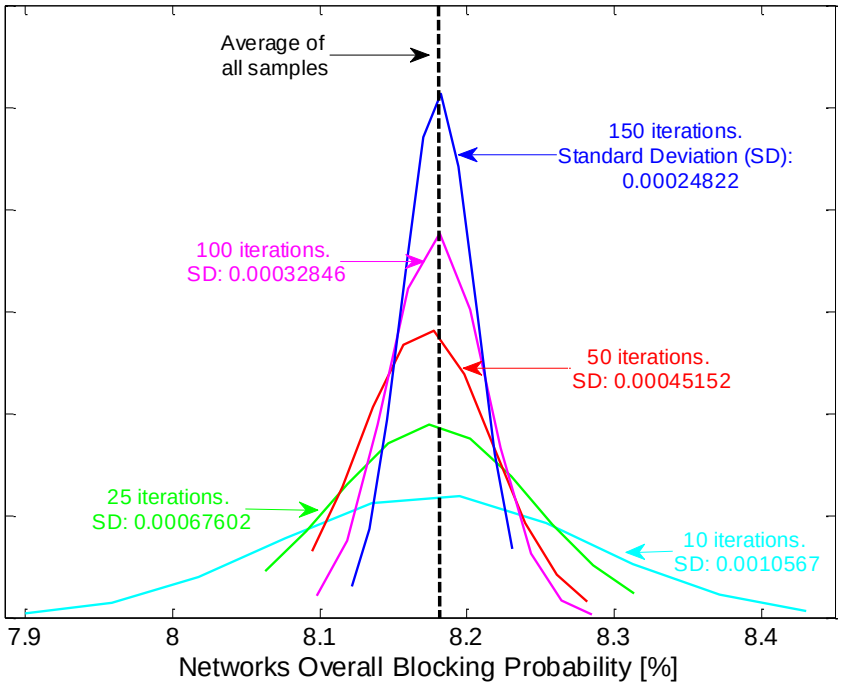


Figure 3.16 Overall call blocking probability Vs Number of iterations (100 samples)

In the comparison of Figure 3.16, there are 1200 active users in networks, and simulation parameters given in Table 3.1 are used. The simulation has been run in 100 computers to collect 100 samples. The data is analysed for 10, 25, 50, 100 and 150 numbers of simulation iterations. The dotted line shows the average value of all samples. The solid curves show the normal cumulative distribution curves from different data sets. The standard deviation (SD)

for 100 iterations is much smaller than that for 10, 25, 50 iterations, which reflects that 100 iterations could reach more accuracy result than 10, 25, 50 iterations. Besides, the standard deviation for 100 iterations approaches that for 150 iterations.

Figure 3.17 further evaluates the effect of number of iterations. In the comparison of Figure 3.17, the simulation result is from one computer (one sample). With the increase of number of iterations, the data gradually approaches 8.177%, which is the mean value of all samples. After 100 iterations, the data is slightly different from 8.177%. Therefore, 100 iterations are reasonable.

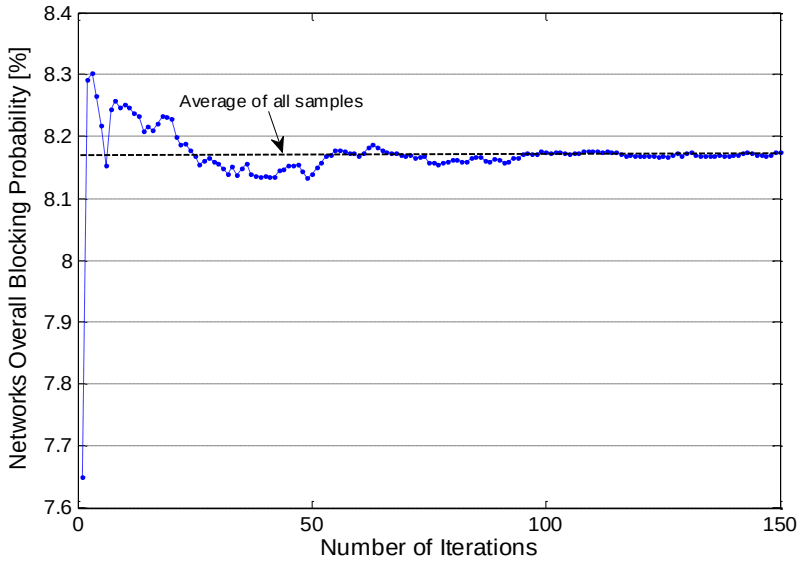


Figure 3.17 Overall call blocking probability Vs Number of iterations (1 sample)

3.9 Summary

In this chapter, a description of the OFDMA system level simulation platform is given. This simulation platform includes three system models, including single-hop cellular networks, user relay cellular networks, and the fixed relay cellular networks. Functionalities of key simulation modules, such as the cell initialisation module, the user distribution module, the channel module and the scheduling module are presented in details. Assumptions and simulation parameters are also given in this chapter.

Chapter 4 Self-Organising Cluster-Based Cooperative Load Balancing

4.1 Introduction

MLB is an effective method to redistribute the traffic load among cells. In this thesis, a self-organising cluster-based cooperative load balancing (CCLB) scheme is proposed. The CCLB scheme aims at overcoming the problems confronted in conventional MLB and improving the load balancing performance.

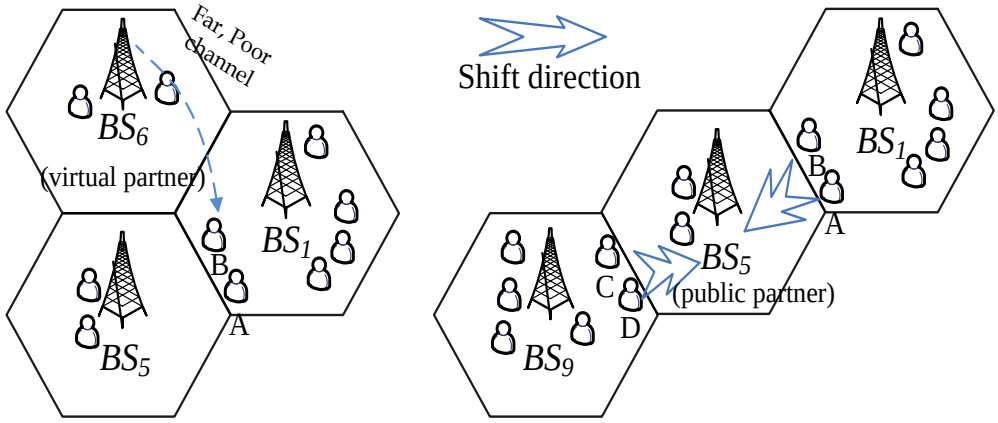
The CCLB scheme is composed of a load balancing clustering stage and a cooperative traffic shifting stage. In the load balancing clustering stage, a user-vote assisted clustering algorithm is designed to select suitable partner cells and avoid the virtual partner problem. In the cooperative traffic shifting stage, both inter-cluster and intra-cluster cooperation are developed. A relative load response model is designed as the inter-cluster cooperation mechanism to mitigate the aggravating load problem. Within each cluster, a traffic offloading optimisation algorithm is designed to reduce the hot-spot cell's load with the cooperation of partners, and minimise partners' average call blocking probability.

Simulation results show that the user-vote assisted clustering algorithm can effectively overcome the virtual partner problem in terms of decreased number of handover offset adjustments and reduced call blocking probability. Simulation results demonstrate that the relative load response model can mitigate the heavily loaded public partner. The effectiveness of the traffic offloading optimisation algorithm is both theoretically analysis and validated by simulation. Results show that the performance of the proposed CCLB

scheme outperforms the conventional MLB.

4.2 Problem Formulation and Challenge

As introduced in Chapter 2, the basic idea of MLB is that the hot-spot cell selects lightly loaded neighbouring cells as partners. Then the hot-spot cell calculates the amount of the required shifting traffic and adjusts HO_{off} towards each partner, and then the cell edge users will be handed over to partners.



(a) Virtual partner problem (b) Aggravating load problem

Figure 4.1 Problems experienced in conventional MLB

4.2.1 Virtual Partner Problem

In conventional MLB, the neighbouring cell's load is widely used as the criterion for finding partner cells. As discussed in Section 2.4.4.1, load based partner selection cannot select the best partner and can lead to the virtual partner problem as shown in Figure 4.1(a). In this thesis, virtual partner is defined as a lightly loaded neighbouring cell, which is far from the heavily loaded cell's edge users. This is because that neighbouring cells with similar load may have different capabilities of serving the shifted users, due to users' random physical position. In order to deal with this problem and select suitable partners, a user-vote assisted clustering algorithm is proposed in Section 4.4.

4.2.2 Public Partner and Aggravating Load Problem

Multiple hot-spot BSs' shifting traffic may result in the aggravating load problem of public

partner. As shown in Figure 4.1(b), BS_5 is the public partner of both hot-spot BS_1 and BS_9. The amount of shifting traffic from each BS is moderate, while the total traffic from two BSs can result in heavily loaded BS_5. In this thesis, the *aggravating load problem* means that the target node, which may be a public partner cell or a relay station (RS), becomes heavily loaded after traffic shifting.

As discussed in Section 2.4.4.2, many conventional MLB schemes ignore the coordination among hot-spot cells, thus they cannot deal with this problem effectively. After analysing the aggravating load problem, a major reason is that a hot-spot cell cannot control other hot-spot cell's shifting traffic to the public partner, under distributed control networks. In order to deal with this problem, a relative load response model, whose basic idea is that the public partner reports the relative load to coordinate multiple hot-spot cells' shifting traffic, is proposed in Section 4.5.1.

4.2.3 Call Blocking Probabilities Increase of Partners

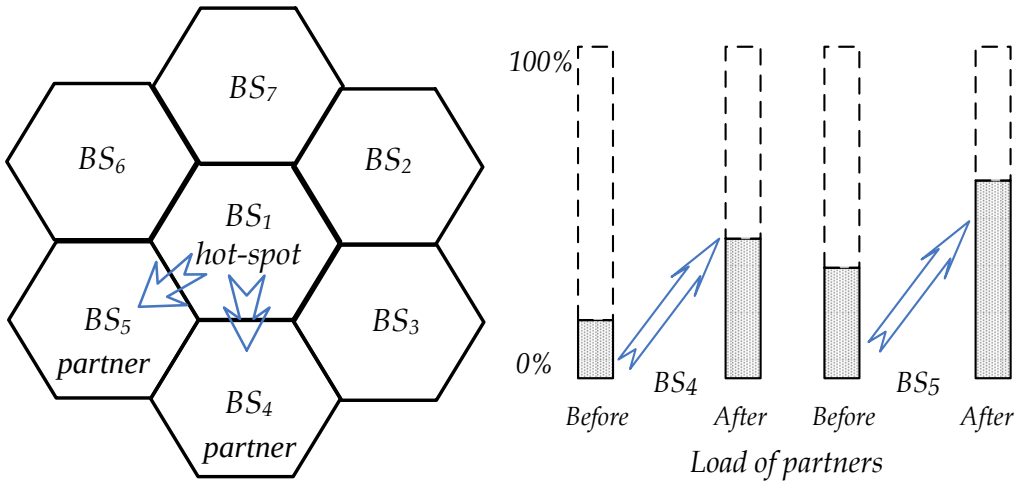


Figure 4.2 Illustration of partners' load increase

As illustrated in Figure 4.2, the shifting traffic from the hot-spot BS_1 will increase the load of partner BS_4 and partner BS_5. From the Erlang loss model [Goldsmith05] [WZ05], the increased load of these partners will increase their call blocking probabilities. An effective MLB scheme should keep the call blocking probabilities of partners at a low level, via shifting suitable traffic to each partner. In Section 4.5.2, a traffic offloading optimisation algorithm is designed to shift hot-spot cell's traffic, in order to minimise partners' average call blocking probability.

4.3 Proposed CCLB scheme

The flowchart of the CCLB scheme is shown in Figure 4.3. The aim of the proposed CCLB scheme is to redistribute the traffic load among cells, and address the virtual partner problem and the aggravating load problem.

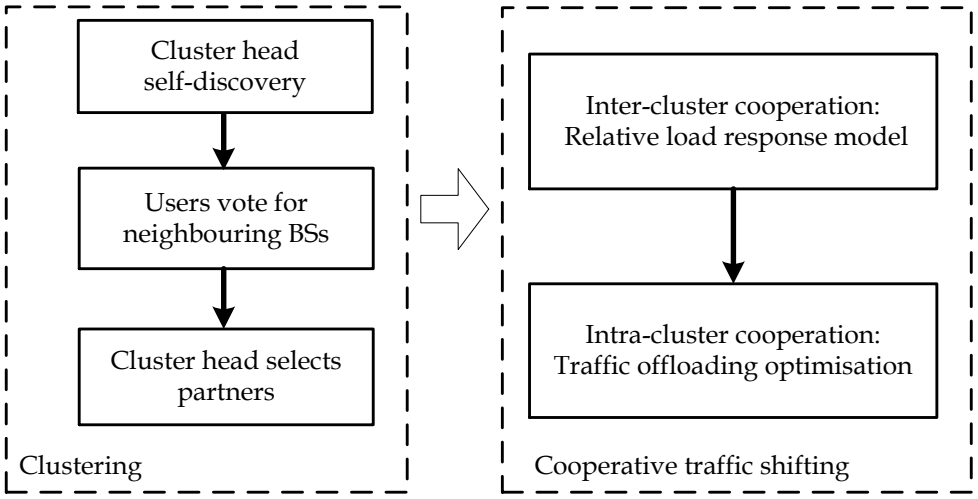


Figure 4.3 Flowchart of CCLB scheme

In the load balancing clustering stage, a hot-spot cell identifies itself as a cluster head by measuring its load condition. Then, the cluster head employs the user-vote model to consider its users' channel condition provided by neighbouring cells. Based on both user-vote model and neighbouring cell's load, the cluster head selects suitable partners to construct load balancing cluster. Compared with the pure load based partner selection mechanism adopted by conventional MLB, the added user-vote model can effectively avoid the virtual partner problem.

The inter-cluster cooperation is designed to overcome the public partner's aggravating load problem. Once a cell is selected by two or more cluster heads, this public partner analyses the traffic shifting requests from all cluster heads and responds with a cluster-specific relative load back to each cluster head. The relative load of a public partner is higher than its actual load. Besides, the relative load is specific to different cluster heads. Based on the public partner's relative load, each cluster head calculates the amount of traffic which can be shifted. Within each cluster, the cluster head employs the Lagrange multiplier method to optimise its shifting traffic to each partner, in order to minimise its partners' average call blocking probability.

The proposed CCLB scheme has the self-organising feature. In the load balancing clustering stage, a hot-spot cell self-discovers as a cluster head and selects its partner cells via cell to cell communication. In addition, this stage is adaptive to time-varying channel condition and load distribution. In the cooperative traffic shifting stage, cluster heads and partners self-optimise the shifting traffic, which is also an essential feature of self-organising networks [E³08].

4.3.1 Cluster Structure

An example of the load balancing cluster structure is illustrated in Figure 4.4. The OFDMA cellular networks suffer an uneven load distribution, and two clusters are constructed for load balancing. The hot-spot cell is defined as the *cluster head*, and *partners* are a subset of neighbouring cells where the cluster head intends to offload traffic. Partners can be classified into two types: a public partner (PP) receives traffic from multiple hot-spot cells; a non-public partner (NP) receives traffic from only one hot-spot cell. Therefore, each load balancing cluster is composed of one cluster head and one or more public partners / non-public partners.

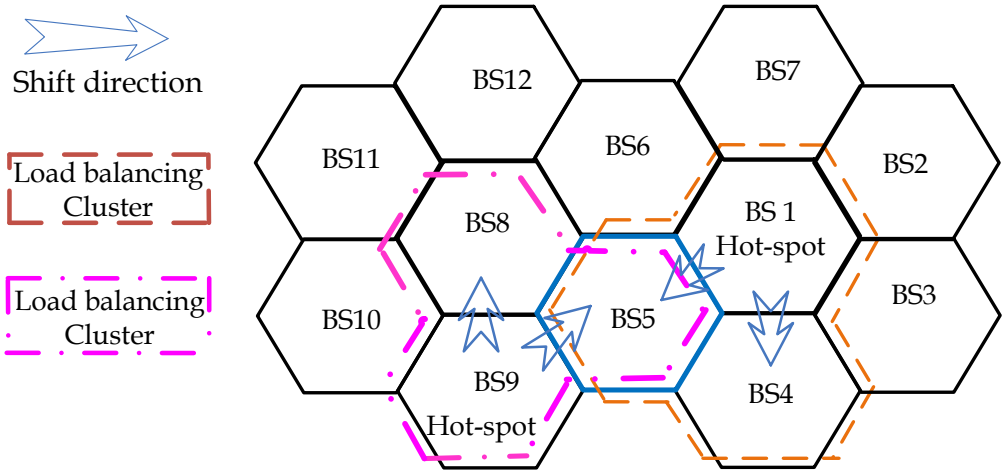


Figure 4.4. Example of two load balancing clusters

For a more general system model, it is assumed that the hot-spot BS_h has I neighbouring BSs and BS_h serves K active users. After the load balancing clustering stage, there are H cluster heads, which are BS_h and BS_j ($j \in \{1,2\ldots H\}, j \neq h$), requesting to shift their traffic to the *public partner* p. In addition, the cluster head BS_h has N non-public partners indexed with n ($n \in \{1,2\ldots N\}$), and P public partners indexed with p ($p \in \{1,2\ldots P\}$). The structure and notation

of load balancing clusters are shown in Figure 4.5.

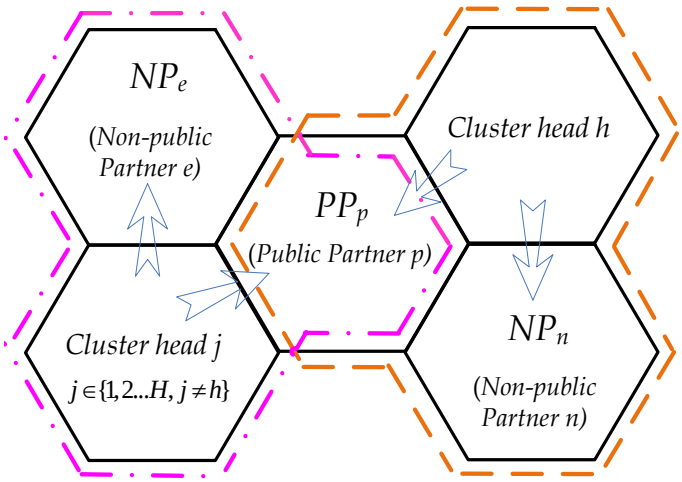


Figure 4.5 Structure and notation of load balancing clusters

4.3.2 Definitions and System Parameters

The definitions and system parameters that will be used in Chapter 4 are listed as follows:

M: Total number of subcarriers in each cell.

M_h: Mean number of subcarriers in use in BS_h, during the load measurement period.

ΔM_h: BS_h tries to release ΔM_h subcarriers by traffic shifting.

L: Each BS's actual load. L is defined as the ratio of the number of subcarriers in use to its total number of subcarriers M, $0\% \leq L \leq 100\%$, e.g., the actual load of BS_h $L_h = M_h / M$ [3GPP10b].

L_{HL}: Threshold of heavy load / hot-spot. A BS is heavy load / hot-spot when this BS's actual load goes above L_{HL}. (L_{HL}=70% [SOCRATES10]. Under 25 physical resource blocks, the call blocking probability of 70%×25Erlang is 2%, based on Erlang loss model [Goldsmith05].)

BS_i: Neighbouring BS_i. Assuming BS_h has I neighbours indexed with i ($i \in \{1 \ldots I\}$).

U_k: *User k.* Assuming BS_h has K active users indexed with k ($k \in \{1 \ldots K\}$).

$SINR_{k,i}^{est}$: U_k's SINR (signal to interference plus noise ratio) estimation towards BS_i.

$SINR_{k,h}$: U_k's SINR in BS_h.

$V_{k,i}$: *Vote* of U_k towards neighbouring BS_i.

Pr_i : The selection priority of neighbouring BS_i in user-vote assisted clustering.

p : Index of public partners, $p \in \{1 \ldots P\}$. PP_p denotes *public partner p*.

n : Index of non-public partners, $n \in \{1 \ldots N\}$. NP_n denotes *non-public partner n*.

h, j : Index of cluster heads. It is assumed that PP_p receives traffic from H cluster heads, consisting of BS_h and BS_j ($j \in \{1,2 \ldots H\}$, $j \neq h$).

L_n : Actual load of NP_n (*non-public partner n*).

L_p : Actual load of PP_p (*public partner p*).

$R_{p,h}$: PP_p's relative load corresponding to the cluster head BS_h.

$\tilde{R}_{p,h}$: PP_p's relative load, after receiving the traffic from the cluster head BS_h. $\tilde{R}_{p,h}$ equals the sum of PP_p's relative load $R_{p,h}$ and its receiving traffic from BS_h.

$\tilde{L}_n$: NP_n's actual load, after receiving traffic from BS_h.

$\tilde{L}_{pars}$: Average load of BS_h's partners, after receiving traffic from BS_h.

$\tilde{B}_n$: NP_n's call blocking probability, after receiving traffic from BS_h.

$\tilde{B}_{p,h}$: PP_p's call blocking probability, after receiving traffic from BS_h.

$\tilde{B}_{pars}$: Average call blocking probability of BS_h's partners, after receiving traffic from BS_h.

M_p^{LBthr} : The receiving traffic threshold of PP_p.

$M_{p,h}^{LB}$: PP_p's subcarriers for receiving BS_h's traffic.

4.4 User-Vote assisted Clustering

The user-vote assisted clustering mechanism is designed to address the virtual partner problem so that cluster head can select suitable partners for effective load balancing.

The proposed clustering algorithm follows three steps. First, a hot-spot BS identifies itself as a cluster head to trigger its load balancing cluster construction. Second, users served by the cluster head estimate their SINR/s provided by all neighbouring BSs, and quantify the value to calculate their votes and report them back to the cluster head. Finally, the cluster head jointly considers the received user-votes and neighbouring BSs' load to effectively select partner cells to construct the load balancing cluster.

4.4.1 User-Vote Model

Figure 4.6 depicts the cluster head self-discovery mechanism. BS_h discovers itself as a cluster head if the period, when its actual load is higher than the threshold L_{HL}, is larger than the critical time T_{crit}. T_{crit} provides hysteresis and helps avoid an incorrect cluster head diagnosis, triggering the cluster construction. T_{crit}=5000ms [RH12].

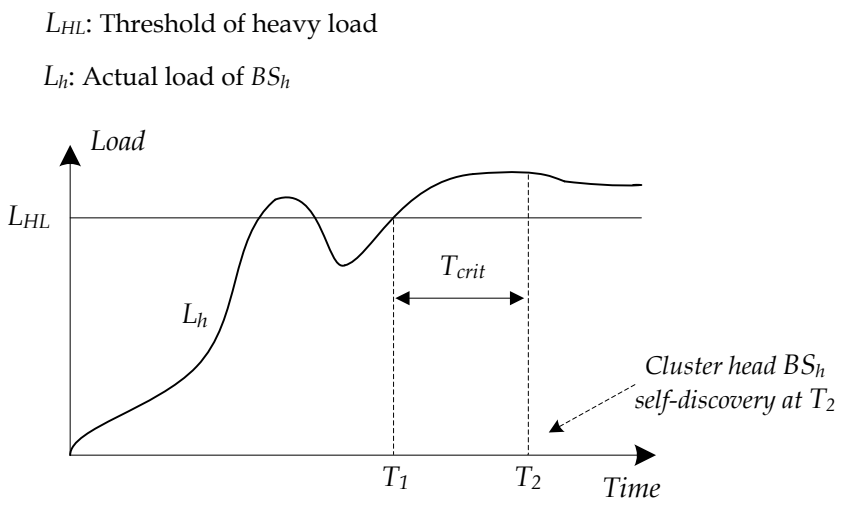


Figure 4.6 Illustration of cluster head self-discovery

The basic idea of the user-vote model is illustrated in Figure 4.7. It is assumed that the cluster head BS_h has I neighbouring BSs indexed with i ($i \in \{1 \ldots I\}$), and BS_h has K active users indexed with k ($k \in \{1 \ldots K\}$). U_k estimates its SINR from neighbouring BS_i as $SINR_{k,i}^{est}$. Based on $SINR_{k,i}^{est}$ and the received $SINR_{k,h}$ from the cluster head, U_k calculates its *vote* of neighbouring BS_i, as $V_{k,i}$. Since U_k is near to two neighbouring BSs at most, U_k reports two neighbouring BSs with the largest non-zero *vote* $V_{k,i}$ to the cluster head.

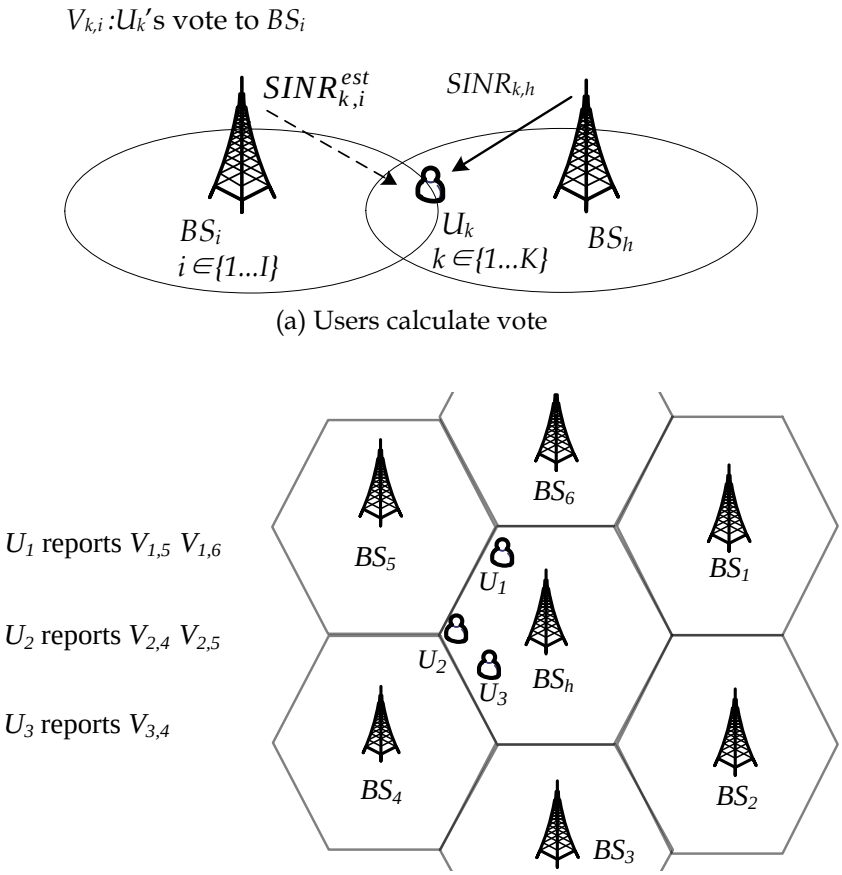


Figure 4.7 User-vote model

1) *SINR Estimation*

U_k estimates its worst SINR from neighbouring BS_i, based on the reference signal received power (RSRP). The precise SINR estimation is difficult because U_k's subcarriers allocated by BS_i vary in both time and frequency domains, based on the service requirements and channel condition of all serving users in BS_i. In OFDMA networks with full frequency reuse [3GPP10c] [HGV10] [RY10], all neighbouring BSs are likely to use the co-channel subcarriers of U_k for transmission at the same time, which induces the inter-cell interference. Therefore, U_k estimates the worst SINR from BS_i using (4.1). $SINR_{k,i}^{est}$ reflects the channel condition after U_k is shifted, and is used to calculate U_k's *vote*.

$$SINR_{k,i}^{est} = \frac{RSRP_{k,i}}{RSRP_{k,h} + \sum_{\bar{i}=1,\bar{i}\neq i}^{I} RSRP_{k,\bar{i}}} \tag{4.1}$$

where $\sum_{\bar{i}=1,\bar{i}\neq i}^{I} RSRP_{k,\bar{i}}$ is from other neighbouring BSs. $RSRP_{k,i}$ and $RSRP_{k,h}$ are from the voting target BS_i and the cluster head BS_h, respectively. In (4.1), the noise is negligible compared with the interference in the full frequency reuse OFDMA cellular networks [3GPP10c] [RY10].

2) *Vote Calculation and Vote Report*

As shown in Figure 4.7(a), U_k quantified $SINR_{k,i}^{est}$ into *vote* as $V_{k,i}$, via the comparison between $SINR_{k,i}^{est}$ and the serving $SINR_{k,h}$ from BS_h. $V_{k,i}$ indicates U_k's probability of being offloaded to neighbouring BS_i, reflecting its satisfaction degree to BS_i.

$$V_{k,i} = \begin{cases} 1 & SINR_{k,i}^{est} \geq \dfrac{SINR_{k,h}}{\eta} \\ Q_{step} \times Floor\left(\dfrac{SINR_{k,i}^{est}}{SINR_{k,h}/\eta} \times \dfrac{1}{Q_{step}} + 0.5\right) & SINR_{k,i}^{est} < \dfrac{SINR_{k,h}}{\eta} \end{cases} \quad (4.2)$$

where *Floor-function* $floor(x)$ gives the largest integer value, which is less than or equal to the value of x. η=4 to get a suitable threshold $\frac{SINR_{k,h}}{\eta}$ to assist U_k to calculate its *vote* towards BS_i. This is because that $SINR_{k,i}^{est} \geq \frac{SINR_{k,h}}{4}$ can identify cell edge users. η=4 is a suitable value analysed in Section 4.8, which is derived from that RSRP differential threshold to judge cell edge user is *3dB* [FSCK10] [SKMNT10].

- For the users with $SINR_{k,i}^{est} \geq \frac{SINR_{k,h}}{\eta}$, they are located at the cell edge of BS_h-to-BS_i, and BS_i can serve them with satisfactory data rate. Hence, they vote for BS_i with full vote $V_{k,i}$ =1.
- For the users with $SINR_{k,i}^{est} < \frac{SINR_{k,h}}{\eta}$, $V_{k,i}$ is based on the ratio of $SINR_{k,i}^{est}$ to $\frac{SINR_{k,h}}{\eta}$.

To save the signalling load of reporting vote, in (4.2) $\frac{SINR_{k,i}^{est}}{SINR_{k,h}/\eta}$ is converted to a discrete $V_{k,i}$ value by the quantization step Q_{step} and the *Floor-function* [TC11], e.g., $V_{k,i} \in \{0, 0.1, 0.2 \ldots 0.9, 1.0\}$ under Q_{step}=0.1. Therefore, $V_{k,i}$ requires a smaller code length than reporting the actual value.

After calculating *vote*, U_k only reports its *vote* for the two neighbouring BSs with the largest non-zero $V_{k,i}$ to the cluster head. Since in most cases, U_k is near to two neighbouring BSs at most, the received channel condition from other neighbouring BSs are so small that U_k cannot be shifted to. This is exemplified in Figure 4.7(b). The non-zero constraint avoids the users, which are very near to BS_h, reporting. This can further reduce the signalling load, compared with the mechanism of user reporting votes of all neighbouring BSs.

4.4.2 Partner Selection

Based on the *vote* report of users, the cluster head calculates the total votes of neighbouring BS_i as $\sum_{k=1}^{K} V_{k,i}$. The total votes reflect the traffic shifting capability of BS_i, affected by users' channel condition. The higher the value, the more users can be shifted to BS_i.

The cluster head also considers the actual load, which reflects the amount of idle subcarriers of neighbouring BS_i that can serve the shifted users. The selection priority of neighbouring BS_i is defined as

$$\Pr_i = \frac{\sum_{k=1}^{K} V_{k,i}}{K} + (1 - L_i) \qquad i \in \{1 \ldots I\} \tag{4.3}$$

where L_i is the actual load of BS_i, K is the total number of active users in the cluster head. The denominator K guarantees the range of total votes is from 0 to 1, which is the same as the actual load ($0 \leq L_i \leq 1$). Therefore, the factor of actual load and the factor of users' vote have the same weight in (4.3).

According to the priority of (4.3), under the same number of votes, the neighbouring BS with lower load has higher priority to be selected as a partner. Meanwhile, under the same load, the neighbouring BS with more votes has higher priority.

The cluster head also employs two filters to improve the efficiency of the clustering. The vote filter is to avoid selecting a neighbouring BS, which has no user from the cluster head located at its edge, as shown in (4.4). The load filter is to avoid selecting a heavily loaded BS, as shown in (4.5).

Vote filter: $$\underset{k\in\{1\ldots K\}}{Max}\, V_{k,i} = 1 \qquad i \in \{1\ldots I\} \tag{4.4}$$

Load filter: $$L_i < L_{HL} \qquad i \in \{1\ldots I\} \tag{4.5}$$

where L_{HL} is the threshold of heavy load.

In the last step of the proposed clustering algorithm, the cluster head finds all neighbouring BSs satisfying the Filters (4.4) and (4.5). Then the cluster head sorts these neighbouring BSs in descending order, according to their selection priorities (4.3). It continuously selects the highest priority neighbouring BS as cluster's partner in sequence, until the number of partners in the cluster is larger than the maximum cluster size. (Section 4.6.1 investigates the appropriate cluster size via simulation analysis.) Then, the cluster head sends a request message to the selected neighbouring BSs for cluster construction. Then, each selected neighbouring BS will respond with a confirmation message, which confirms being the partner in the load balancing cluster. This process can be implemented over the X2 interface in LTE [3GPP10a] [3GPP11a]. The clustering algorithm is finished after partners' response.

After the cluster construction, the cluster head shifts traffic to its partners, and this stage is described in Section 4.5. After traffic shifting, the cluster head sends the *leave* request to all partners within its cluster. The load balancing cluster will be dismissed after partners respond to *leave*.

4.4.3 Signalling load and Complexity

The signalling load and the computational complexity of the user-vote assisted clustering algorithm are analysed in this section. Figure 4.8 shows the process of this algorithm.

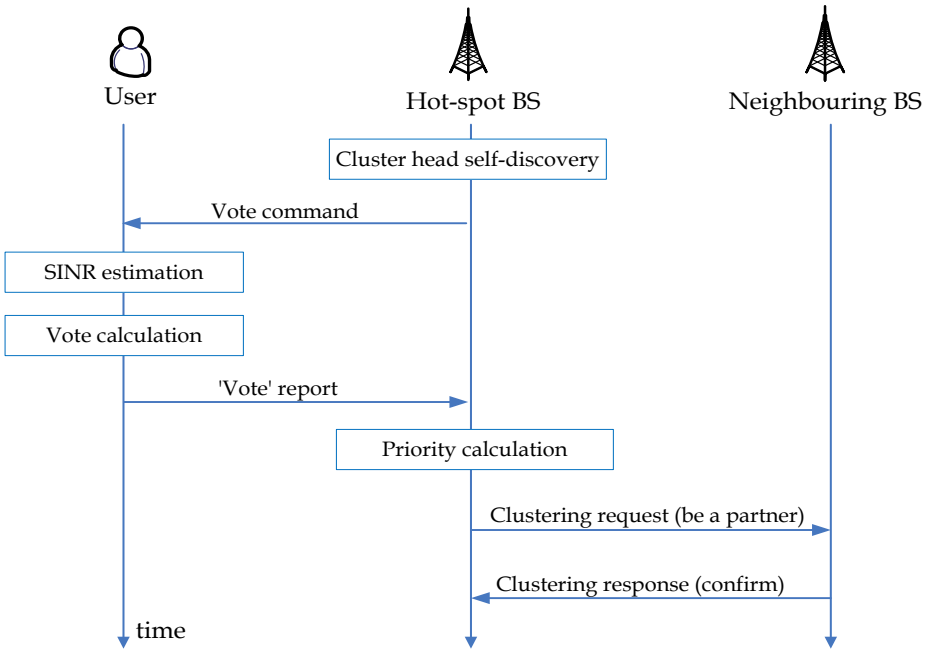


Figure 4.8 Process of user-vote assisted clustering

First, users' SINR estimation is purely based on RSRP, which is available for existing radio resource management (RRM) functionality without extra measurements. Second, in order to save the signalling load of 'User-to-Cluster head' link resulted from the vote report process, each user only reports its *vote* of two neighbouring BSs with the largest non-zero *vote*, rather than reporting all neighbouring BSs' *vote*. Third, the cluster head sends/responds clustering request with partners via cell-to-cell communication. This process consumes the similar signalling load as the partner request/response process in conventional MLB schemes, such as [NA07] [KAPTK10] [LLZL10] [LSJB10] [ZQMZ10a] [YLCW12].

The complexity of each user calculating the *vote* of all neighbouring BSs is $I \times O(I)$, the complexity of reporting two largest $V_{k,i}$ neighbouring BSs is $O(2I\text{-}3)$. Hence, the complexity of the user-vote model is $K \times I \times O(I) + K \times O(2I\text{-}3)$. In the partner selection step, the complexity of priority calculation and the filters of each neighbouring BS are $O(K)$ and $O(K)+O(1)$, respectively. Hence, the complexity of the partner selection is $I \times 2 \times O(K) + I \times O(1)$. Therefore, its overall complexity is $K \times I \times O(I) + K \times O(2I\text{-}3) + 2I \times O(K) + I \times O(1)$.

4.5 Cooperative Traffic Shifting

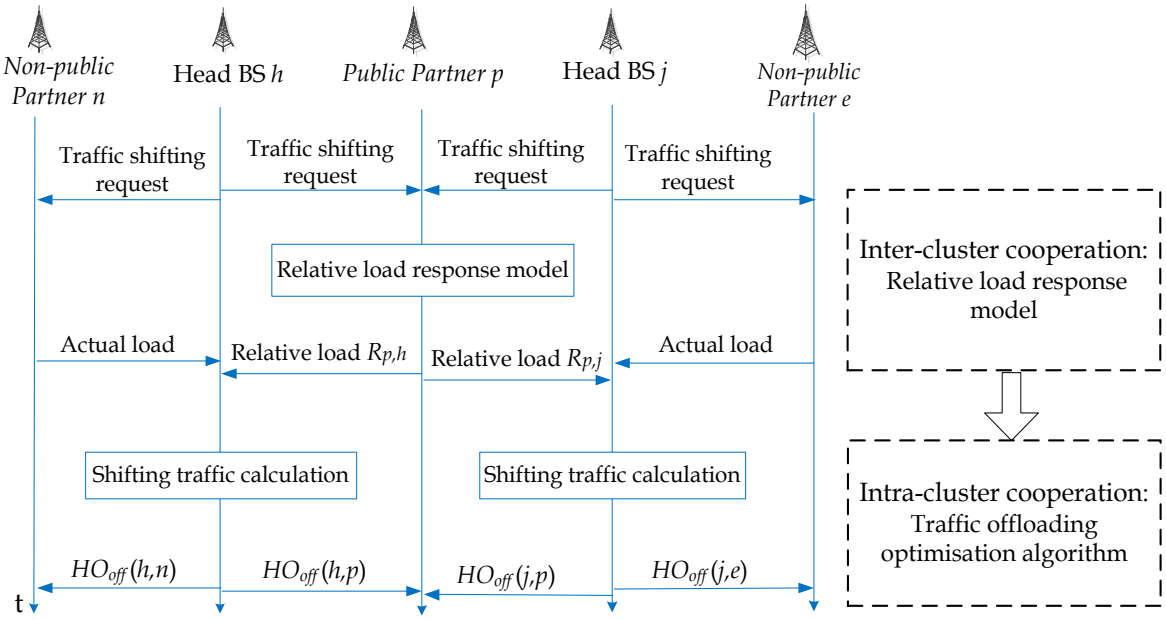


Figure 4.9 Process of cooperative traffic shifting

Once the construction of the load balancing cluster is completed, the cluster head is associated with one or more partner cells. This section presents the proposed cooperative traffic shifting algorithm. Its aim is effectively shifting the cluster head's traffic to address the aggravating load problem, as well as minimising the partners' average call blocking probability in a cluster. Figure 4.9 shows its process under the clusters structure of Figure 4.5. The *public partner* analyses traffic shifting requests of multiple cluster heads, and replies to each cluster head with its cluster-specific relative load, thus addressing the aggravating load problem. Meanwhile, the *non-public partner* replies its actual load to the dedicated cluster head. Then, the cluster head employs the traffic offloading optimisation algorithm to calculate the amount of shifting traffic to each partner, in order to minimise partners' average call blocking probability. Based on the amount of the required shifting traffic, the cluster head adjusts HO_{off} to offload cell edge users.

4.5.1 Inter-Cluster Cooperation: Relative Load Response Model

In relative load response model (RLRM), the public partner analyses its threshold of idle spectrum for receiving traffic. Then it pre-allocates the idle spectrum to each cluster head's shifting traffic. Finally, the public partner calculates its cluster-specific relative load and

reports it back to the corresponding cluster head. According to the received relative load, each cluster can shift a proper amount of traffic to the public partner, thus addressing the aggravating load problem.

4.5.1.1 Public Partner's Load Balancing Spectrum Analysis

From the system model, *public partner p* (PP_p) receives traffic shifting requests from H different cluster heads. In order to use PP_p's idle spectrum to balance multiple cluster heads' load and to avoid PP_p being heavily loaded, Equation (4.6a) shows that PP_p's receiving traffic $M_p^{LB} \le (L_{HL} - L_p) \times M$. Then, PP_p calculates its receiving traffic threshold M_p^{LBthr}, as (4.6b).

$$L_p + \frac{M_p^{LB}}{M} \le L_{HL} \Rightarrow M_p^{LB} \le (L_{HL} - L_p) \times M \tag{4.6a}$$

$$\Rightarrow M_p^{LBthr} = (L_{HL} - L_p) \times M \tag{4.6b}$$

where L_p is the actual load of PP_p, L_{HL} is the threshold of heavy load, M is the total number of subcarriers in each cell. According to Equation (4.6a) and (4.6b), PP_p*'s* subcarriers for serving all cluster heads' users cannot exceed M_p^{LBthr}, otherwise PP_p will become heavily loaded. Then RLRM pre-allocates these M_p^{LBthr} subcarriers to cluster heads.

The system model shows that these H cluster heads consist of BS_h and BS_j ($j \in \{1,2 \ldots H\}$, $j \neq h$). The process of load balancing subcarriers analysis is illustrated in Figure 4.10. PP_p pre-allocates these M_p^{LBthr} subcarriers into two parts: the load balancing subcarriers for receiving traffic from BS_h as $M_{p,h}^{LB}$; the load balancing subcarriers for receiving traffic from BS_j ($j \in \{1 \ldots H\}$, $j \neq h$) as $\sum_{j=1, j\neq h}^{H} M_{p,j}^{LB}$. PP_p pre-allocates more load balancing subcarriers to a higher-loaded cluster head. Hence, $M_{p,h}^{LB}$ is calculated based on BS_h's actual load L_h, using

$$M_{p,h}^{LB} = M_p^{LBthr} \times \frac{L_h}{L_h + \sum_{\breve{j}=1, \breve{j}\neq h}^{H} L_{\breve{j}}} = \frac{L_h \times M \times (L_{HL} - L_p)}{L_h + \sum_{\breve{j}=1, \breve{j}\neq h}^{H} L_{\breve{j}}} \tag{4.7}$$

The amount of shifting traffic from BS_h cannot exceed $M_{p,h}^{LB}$. Similarly, PP_p's load balancing subcarriers for BS_j, $\sum_{j=1, j\neq h}^{H} M_{p,j}^{LB}$ is calculated based on BS_j's actual load L_j, using (4.8).

$$\sum_{\substack{j=1\\ j\neq h}}^{H} M_{p,j}^{LB} = \sum_{\substack{j=1\\ j\neq h}}^{H} \frac{M_p^{LBthr} \times L_j}{L_h + \sum_{\breve{j}=1,\breve{j}\neq h}^{H} L_{\breve{j}}} = \sum_{\substack{j=1\\ j\neq h}}^{H} \frac{L_j \times M \times (L_{HL} - L_p)}{L_h + \sum_{\breve{j}=1,\breve{j}\neq h}^{H} L_{\breve{j}}} \tag{4.8}$$

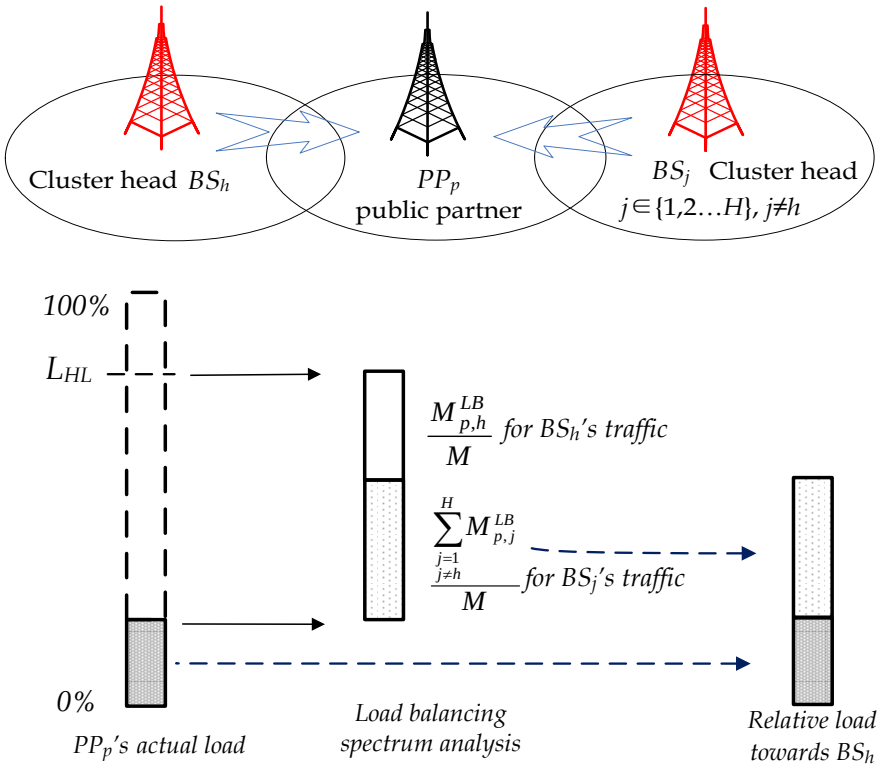


Figure 4.10 Public partner's load balancing spectrum analysis

4.5.1.2 Cluster-Specific Relative Load

From PP_p's actual load $L_p = M_p / M$, M_p subcarriers are used by PP_p itself. Hence, BS_h's shifted users cannot use both M_p and $\sum_{j=1,j\neq h}^{H} M_{p,j}^{LB}$ ($\sum_{j=1,j\neq h}^{H} M_{p,j}^{LB}$ is PP_p's load balancing subcarriers for BS_j's traffic). Therefore, PP_p calculates its cluster-specific relative load towards BS_h as $R_{p,h}$, using (4.9a). In (4.9b), the relative load is converted to a discrete value by the quantization step Q_s and the *Floor-function* [TC11], e.g., $R_{p,h} \in \{0, 0.01, 0.02 \ldots 0.99, 1.0\}$ under $Q_s = 0.01$. Finally, PP_p informs BS_h with $R_{p,h}$ as the response.

$$R_{p,h} \Leftarrow \left(\sum_{\substack{j=1\\ j\neq h}}^{H} M_{p,j}^{LB} / M\right) + L_p \tag{4.9a}$$

$$\Rightarrow \quad R_{p,h} = Q_s \times Floor\left\{\left[\sum_{\substack{j=1\\ j\neq h}}^{H} \frac{L_j (L_{HL} - L_p)}{(L_h + \sum_{\breve{j}=1,\breve{j}\neq h}^{H} L_{\breve{j}})} + L_p\right] \frac{1}{Q_s} + 0.5\right\} \tag{4.9b}$$

$R_{p,h}$ reflects PP_p's traffic shifting capability towards BS_h. The capability is decided by both

PP_p's idle spectrum and all clusters' traffic shifting requests. After the relative load response, each cluster head can estimate its maximum amount of shifting traffic to PP_p, in order to shift an appropriate amount of traffic. In addition, the relative load of the public partner is always higher than its actual load, and hence the cluster head can shift more traffic to other non-public partners and less traffic to the public partner. Therefore, RLRM assists the public partner to coordinate multiple clusters' traffic shifting, and address the heavily loaded public partner.

4.5.2 Intra-Cluster Cooperation: Traffic Offloading Optimisation Algorithm

In the traffic shifting stage, different load balancing schemes have different load reduction objectives for the cluster head BS_h [DSJ97] [SV09] [SOCRATES10] [WTJLHL10]. For example, some schemes try to reduce the hot-spot cell's load to a pre-defined threshold, while some other schemes try to reduce the hot-spot cell's load to its neighbouring cells' average load. In order to design a load balancing scheme to meet different load reduction requirements, this scheme does not pre-define the load reduction value/threshold. Instead, the proposed scheme assumes BS_h tries to release ΔM_h subcarriers, which has different values according to different load reduction objectives.

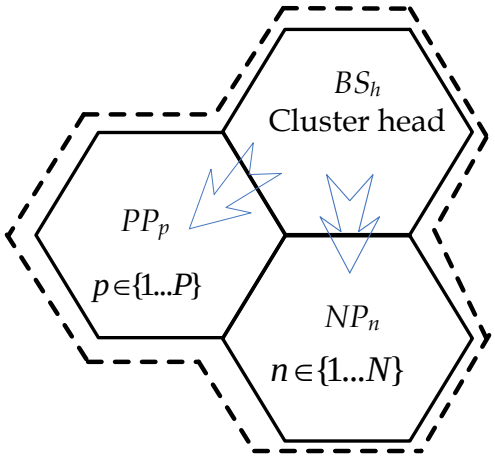


Figure 4.11 Cluster model of BS_h

Figure 4.11 is the cluster model of BS_h introduced in Section 4.3. Assuming BS_h has N non-public partners denoted as NP_n $n \in \{1,2...N\}$, and P public partners denoted as PP_p $p \in \{1,2...P\}$. Since BS_h tries to offload ΔM_h traffic to its partners, its load reduction $\Delta L_h = \Delta M_h / M$. This will increase its partners' load and call blocking probability. The load and call blocking

probability of BS_h's partners are listed as follows:

- Initial load: $L_1 \ldots L_n \ldots L_N$ of non-public partners; $R_{1,h} \ldots R_{p,h} \ldots R_{P,h}$ of public partners.
- Load after receiving BS_h traffic: $\tilde{L}_1 \ldots \tilde{L}_n \ldots \tilde{L}_N$ of non-public partners; $\tilde{R}_{1,h} \ldots \tilde{R}_{p,h} \ldots \tilde{R}_{P,h}$ of public partners.
- Call blocking probability after receiving BS_h traffic: $\tilde{B}_1 \ldots \tilde{B}_n \ldots \tilde{B}_N$ of non-public partners; $\tilde{R}_{1,h} \ldots \tilde{R}_{p,h} \ldots \tilde{R}_{P,h}$ of public partners.

Therefore, the traffic offloading optimisation algorithm controls BS_h's shifting traffic to each partner, in order to minimise these partners' average call blocking probability.

4.5.2.1 Objective: Minimise Partners' Average Blocking Probability

After receiving traffic from BS_h, PP_p's relative load is denoted as $\tilde{R}_{p,h}$, which equals the sum of its relative load $R_{p,h}$ and BS_h's shifting traffic. After traffic shifting, NP_n's actual load is denoted as $\tilde{L}_n$, which equals the sum of its actual load L_n and BS_h's shifting traffic. Therefore, under the cluster head's load reduction ΔL_h, all partners' total load is expressed as

$$\tilde{L}_{pars}^{all} = \Delta L_h + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h} = \sum_{n=1}^{N} \tilde{L}_n + \sum_{p=1}^{P} \tilde{R}_{p,h} \tag{4.10}$$

The Erlang loss model is widely used to evaluate the grade of service (GoS) in wireless networks [Goldsimith05] [WZ05]. After receiving traffic from BS_h, the call blocking probability of NP_n and PP_p are calculated based on the Erlang loss model, as $\tilde{B}_n$ in Equation (4.11a) and $\tilde{B}_{p,h}$ in Equation (4.11b), respectively.

$$\tilde{B}_n = \frac{(\tilde{L}_n \times M)^M / M!}{\sum_{k=0}^{M} (\tilde{L}_n \times M)^k / k!} \qquad n \in \{1 \ldots . N\} \tag{4.11a}$$

$$\tilde{B}_{p,h} = \frac{(\tilde{R}_{p,h} \times M)^M / M!}{\sum_{k=0}^{M} (\tilde{R}_{p,h} \times M)^k / k!} \qquad p \in \{1 \ldots . P\} \tag{4.11b}$$

According to the formula of average call blocking probability in a system [ZY91] [GKTH97], the average call blocking probability of BS_h's partners is defined as

$$\tilde{B}_{pars} = \frac{\sum_{n=1}^{N} \tilde{B}_n \tilde{L}_n + \sum_{p=1}^{P} \tilde{B}_{p,h} \tilde{R}_{p,h}}{\tilde{L}_{pars}^{all}} \tag{4.12}$$

Under BS_h's load reduction ΔL_h, the optimisation objective of minimising its partners' average call blocking probability $\tilde{B}_{pars}$, is formulated as (4.13)-(4.16).

$$\underset{\tilde{L}_n, \tilde{R}_{p,h}}{MIN} \tilde{B}_{pars} = \underset{\tilde{L}_n, \tilde{R}_{p,h}}{MIN} \frac{\sum_{n=1}^{N} \tilde{B}_n \tilde{L}_n + \sum_{p=1}^{P} \tilde{B}_{p,h} \tilde{R}_{p,h}}{\tilde{L}_{pars}^{all}} \tag{4.13}$$

Subject to $$\tilde{L}_{pars}^{all} - \sum_{n=1}^{N} \tilde{L}_n - \sum_{p=1}^{P} \tilde{R}_{p,h} = 0 \tag{4.14}$$

$$\tilde{L}_n > L_n \Rightarrow -(\tilde{L}_n - L_n) < 0 \Rightarrow \tilde{L}_n - L_n > 0 \qquad n \in \{1..N\} \tag{4.15}$$

$$\tilde{R}_{p,h} > R_{p,h} \Rightarrow -(\tilde{R}_{p,h} - R_{p,h}) < 0 \Rightarrow \tilde{R}_{p,h} - R_{p,h} > 0 \quad p \in \{1..P\} \tag{4.16}$$

The total load constraint of (4.14) is derived from Equation (4.10). Since each non-public partner receives traffic from the cluster head, this will increase the actual load of each non-public partner, and this constraint is depicted as (4.15). Similarly, the shifting traffic from the cluster head will increase the relative load of each public partner, and this constraint is depicted as (4.16).

4.5.2.2 Optimisation Method

In order to minimise $\tilde{B}_{pars}$, the Lagrange multipliers method and Karush-Kuhn-Tucker (KKT) condition [Bertsekas99] [Hanson99] are employed. The Lagrange multiplier λ is introduced for the constraint (4.14). In addition, the Lagrange multiplier vectors $\vec{\omega} = \{\omega_1, \omega_2 ... \omega_N\}$ and $\vec{\mu} = \{\mu_1, \mu_2 ... \mu_p\}$ are introduced for the constraints (4.15) and (4.16), respectively.

1) First, the objective formulated in (4.13)-(4.16) is defined as the Lagrangian function

$$F(\tilde{L}_n, \tilde{R}_{p,h}) = \frac{\sum_{n=1}^{N} \tilde{B}_n \tilde{L}_n + \sum_{p=1}^{P} \tilde{B}_{p,h} \tilde{R}_{p,h}}{\tilde{L}_{pars}^{all}} - \lambda(\tilde{L}_{pars}^{all} - \sum_{n=1}^{N} \tilde{L}_n - \sum_{p=1}^{P} \tilde{R}_{p,h}) + \sum_{n=1}^{N} \omega_n \times (-1) \times (\tilde{L}_n - L_n) + \sum_{p=1}^{P} \mu_p \times (-1) \times (\tilde{R}_{p,h} - R_{p,h})$$

$$= \frac{\sum_{n=1}^{N} \tilde{B}_n \tilde{L}_n + \sum_{p=1}^{P} \tilde{B}_{p,h} \tilde{R}_{p,h}}{\tilde{L}_{pars}^{all}} - \lambda(\tilde{L}_{pars}^{all} - \sum_{n=1}^{N} \tilde{L}_n - \sum_{p=1}^{P} \tilde{R}_{p,h}) - \sum_{n=1}^{N} \omega_n \times (\tilde{L}_n - L_n) - \sum_{p=1}^{P} \mu_p \times (\tilde{R}_{p,h} - R_{p,h}) \quad (4.17)$$

where $\tilde{B}_n$ is the function of variable $\tilde{L}_n$ shown in (4.11a), $\tilde{B}_{p,h}$ is the function of variable $\tilde{R}_{p,h}$ shown in (4.11a).

According to the KKT condition, for $n \in \{1,2...N\}$, there is $\omega_n \times (\tilde{L}_n - L_n) = 0$. Meanwhile, (4.15) shows $\tilde{L}_n - L_n > 0$. Therefore, the Lagrange multiplier $\omega_n = 0$ when $n = 1,2...N$.

Similarly, the KKT condition requires $\mu_p \times (\tilde{R}_{p,h} - R_{p,h}) = 0$, and (4.16) has the constraint $\tilde{R}_{p,h} - R_{p,h} > 0$. Therefore, the Lagrange multiplier $\mu_p = 0$ when $p = 1,2...P$.

2) Second, the partial derivative $\frac{\partial F}{\partial \tilde{L}_n}$ $n \in \{1,2...N\}$ and $\frac{\partial F}{\partial \tilde{R}_{p,h}}$ $p \in \{1,2...P\}$ are given by (4.18), (4.19).

$$\frac{\partial F}{\partial \tilde{L}_n} = \frac{\partial(\tilde{B}_n \times \tilde{L}_n)}{\partial \tilde{L}_n} \frac{1}{\tilde{L}_{pars}^{all}} - \lambda \frac{\partial(\tilde{L}_{pars}^{all} - \tilde{L}_n - \sum_{\bar{n}=1,\bar{n}\neq n}^{N} \tilde{L}_{\bar{n}} - \sum_{p=1}^{P} \tilde{R}_{p,h})}{\partial \tilde{L}_n} - 0 - 0 \quad (4.18a)$$

$$\Rightarrow \quad \frac{\partial F}{\partial \tilde{L}_n} = \frac{(M+1)\tilde{B}_n}{\tilde{L}_{pars}^{all}} - \frac{\sum_{k=0}^{M} (\tilde{L}_n \times M)^k \times k/k!}{\sum_{k=0}^{M} (\tilde{L}_n \times M)^k / k!} \frac{\tilde{B}_n}{\tilde{L}_{pars}^{all}} + \lambda \quad (4.18b)$$

$$\frac{\partial F}{\partial \tilde{R}_{p,h}} = \frac{\partial(\tilde{B}_{p,h} \times \tilde{R}_{p,h})}{\partial \tilde{R}_{p,h}} \frac{1}{\tilde{L}_{pars}^{all}} - \lambda \frac{\partial(\tilde{L}_{pars}^{all} - \tilde{R}_{p,h} - \sum_{\bar{p}=1,\bar{p}\neq p}^{P} \tilde{R}_{\bar{p},h} - \sum_{n=1}^{N} \tilde{L}_n)}{\partial \tilde{R}_{p,h}} - 0 - 0 \quad (4.19a)$$

$$\Rightarrow \frac{\partial F}{\partial \tilde{R}_{p,h}} = \frac{(M+1)\tilde{B}_{p,h}}{\tilde{L}_{pars}^{all}} - \frac{\sum_{k=0}^{M} (\tilde{R}_{p,h} \times M)^k \times k/k!}{\sum_{k=0}^{M} (\tilde{R}_{p,h} \times M)^k / k!} \frac{\tilde{B}_{p,h}}{\tilde{L}_{pars}^{all}} + \lambda \quad (4.19b)$$

where $\tilde{B}_n$ function and $\tilde{B}_{p,h}$ functions refer to (4.11a) and (4.11b), respectively. Hence, equations (4.20) is constructed to get the solution to $\frac{\partial F}{\partial \tilde{L}_n}$ $n \in \{1,2...N\}$ and $\frac{\partial F}{\partial \tilde{R}_{p,h}}$ $p \in \{1,2...P\}$.

$$
\begin{cases}
\dfrac{\partial F}{\partial \tilde{L}_1} = \dfrac{(M+1)\tilde{B}_1}{\tilde{L}_{pars}^{all}} - \dfrac{\sum_{k=0}^{M}(\tilde{L}_1 \times M)^k \times k/k!}{\sum_{k=0}^{M}(\tilde{L}_1 \times M)^k/k!}\dfrac{\tilde{B}_1}{\tilde{L}_{pars}^{all}} + \lambda = 0 \\
\qquad \cdots\cdots\cdots\cdots\cdots\cdots \\
\dfrac{\partial F}{\partial \tilde{L}_n} = \dfrac{(M+1)\tilde{B}_n}{\tilde{L}_{pars}^{all}} - \dfrac{\sum_{k=0}^{M}(\tilde{L}_n \times M)^k \times k/k!}{\sum_{k=0}^{M}(\tilde{L}_n \times M)^k/k!}\dfrac{\tilde{B}_n}{\tilde{L}_{pars}^{all}} + \lambda = 0 \\
\dfrac{\partial F}{\partial \tilde{L}_N} = \dfrac{(M+1)\tilde{B}_N}{\tilde{L}_{pars}^{all}} - \dfrac{\sum_{k=0}^{M}(\tilde{L}_N \times M)^k \times k/k!}{\sum_{k=0}^{M}(\tilde{L}_N \times M)^k/k!}\dfrac{\tilde{B}_n}{\tilde{L}_{pars}^{all}} + \lambda = 0 \\
\dfrac{\partial F}{\partial \tilde{R}_{1,h}} = \dfrac{(M+1)\tilde{B}_{1,h}}{\tilde{L}_{pars}^{all}} - \dfrac{\sum_{k=0}^{M}(\tilde{R}_{1,h} \times M)^k \times k/k!}{\sum_{k=0}^{M}(\tilde{R}_{1,h} \times M)^k/k!}\dfrac{\tilde{B}_{1,h}}{\tilde{L}_{pars}^{all}} + \lambda = 0 \\
\qquad \cdots\cdots\cdots\cdots\cdots\cdots \\
\dfrac{\partial F}{\partial \tilde{R}_{p,h}} = \dfrac{(M+1)\tilde{B}_{p,h}}{\tilde{L}_{pars}^{all}} - \dfrac{\sum_{k=0}^{M}(\tilde{R}_{p,h} \times M)^k \times k/k!}{\sum_{k=0}^{M}(\tilde{R}_{p,h} \times M)^k/k!}\dfrac{\tilde{B}_{p,h}}{\tilde{L}_{pars}^{all}} + \lambda = 0 \\
\dfrac{\partial F}{\partial \tilde{R}_{P,h}} = \dfrac{(M+1)\tilde{B}_{P,h}}{\tilde{L}_{pars}^{all}} - \dfrac{\sum_{k=0}^{M}(\tilde{R}_{P,h} \times M)^k \times k/k!}{\sum_{k=0}^{M}(\tilde{R}_{P,h} \times M)^k/k!}\dfrac{\tilde{B}_{P,h}}{\tilde{L}_{pars}^{all}} + \lambda = 0
\end{cases}
\Rightarrow \tag{4.20a}
$$

$$
\begin{cases}
\dfrac{\partial F}{\partial \tilde{L}_1} = \dfrac{\dfrac{(\tilde{L}_1 \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{L}_1 \times M)^k (M+1-k)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{L}_1 \times M)^k/k!]^2} + \lambda = 0 \\
\qquad \cdots\cdots\cdots\cdots \\
\dfrac{\partial F}{\partial \tilde{L}_n} = \dfrac{\dfrac{(\tilde{L}_n \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{L}_n \times M)^k (M+1-k)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{L}_n \times M)^k/k!]^2} + \lambda = 0 \\
\dfrac{\partial F}{\partial \tilde{L}_N} = \dfrac{\dfrac{(\tilde{L}_N \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{L}_N \times M)^k (M+1-k)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{L}_N \times M)^k/k!]^2} + \lambda = 0 \\
\dfrac{\partial F}{\partial \tilde{R}_{1,h}} = \dfrac{\dfrac{(\tilde{R}_{1,h} \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{R}_{1,h} \times M)^k (M+1-k)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{R}_{1,h} \times M)^k/k!]^2} + \lambda = 0 \\
\qquad \cdots\cdots\cdots\cdots \\
\dfrac{\partial F}{\partial \tilde{R}_{p,h}} = \dfrac{\dfrac{(\tilde{R}_{p,h} \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{R}_{p,h} \times M)^k (M+1-k)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{R}_{p,h} \times M)^k/k!]^2} + \lambda = 0 \\
\dfrac{\partial F}{\partial \tilde{R}_{P,h}} = \dfrac{\dfrac{(\tilde{R}_{P,h} \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{R}_{P,h} \times M)^k (M+1-k)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{R}_{P,h} \times M)^k/k!]^2} + \lambda = 0
\end{cases}
\Rightarrow
\begin{cases}
\lambda = \dfrac{\dfrac{(\tilde{L}_1 \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{L}_1 \times M)^k (k-M-1)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{L}_1 \times M)^k/k!]^2} \\
\qquad \cdots\cdots\cdots\cdots \\
\lambda = \dfrac{\dfrac{(\tilde{L}_n \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{L}_n \times M)^k (k-M-1)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{L}_n \times M)^k/k!]^2} \\
\lambda = \dfrac{\dfrac{(\tilde{L}_N \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{L}_N \times M)^k (k-M-1)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{L}_N \times M)^k/k!]^2} \\
\lambda = \dfrac{\dfrac{(\tilde{R}_{1,h} \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{R}_{1,h} \times M)^k (k-M-1)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{R}_{1,h} \times M)^k/k!]^2} \\
\qquad \cdots\cdots\cdots\cdots \\
\lambda = \dfrac{\dfrac{(\tilde{R}_{p,h} \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{R}_{p,h} \times M)^k (k-M-1)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{R}_{p,h} \times M)^k/k!]^2} \\
\lambda = \dfrac{\dfrac{(\tilde{R}_{P,h} \times M)^M}{M!} \times \dfrac{\sum_{k=0}^{M}(\tilde{R}_{P,h} \times M)^k (k-M-1)}{k!}}{\tilde{L}_{pars}^{all}[\sum_{k=0}^{M}(\tilde{R}_{P,h} \times M)^k/k!]^2}
\end{cases}
\tag{4.20b}
$$

3) Third, the solution to Equations (4.20) is that λ is expressed as (4.21), $\tilde{L}_n$ and $\tilde{R}_{p,h}$ are expressed as (4.22).

$$\lambda = \frac{(\frac{\sum_{n=1}^{N}\tilde{L}_n + \sum_{p=1}^{P}\tilde{R}_{p,h}}{N+P} \times M)^M \Big/ M!}{(\sum_{n=1}^{N}\tilde{L}_n + \sum_{p=1}^{P}\tilde{R}_{p,h})} \times \frac{\sum_{k=0}^{M} \frac{(\frac{\sum_{n=1}^{N}\tilde{L}_n + \sum_{p=1}^{P}\tilde{R}_{p,h}}{N+P} \times M)^k \times [k-(M+1)]}{k!}}{\{\sum_{k=0}^{M}(\frac{\sum_{n=1}^{N}\tilde{L}_n + \sum_{p=1}^{P}\tilde{R}_{p,h}}{N+P} \times M)^k / k!\}^2} \tag{4.21}$$

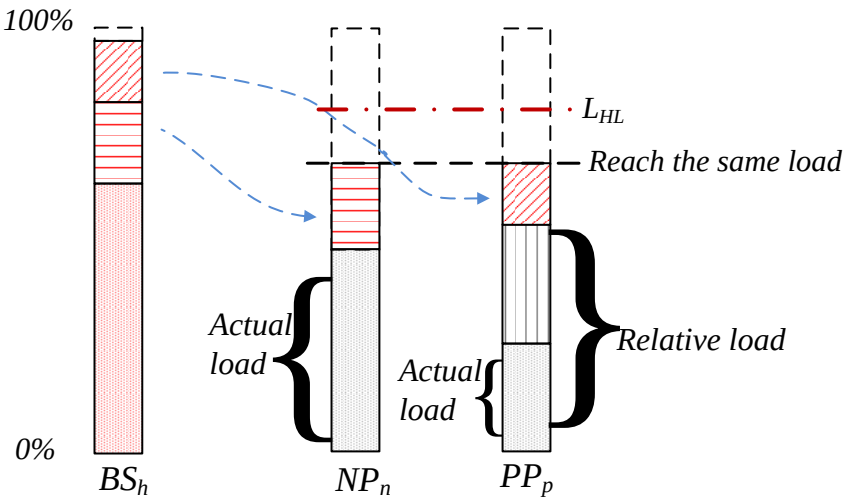
$$\begin{cases} \tilde{L}_n = \dfrac{\sum_{\breve{n}=1}^{N}\tilde{L}_{\breve{n}} + \sum_{\breve{p}=1}^{P}\tilde{R}_{\breve{p},h}}{N+P} & n \in \{1...N\} \\ \tilde{R}_{p,h} == \dfrac{\sum_{\breve{n}=1}^{N}\tilde{L}_{\breve{n}} + \sum_{\breve{p}=1}^{P}\tilde{R}_{\breve{p},h}}{N+P} & p \in \{1...P\} \end{cases} \tag{4.22}$$

The value of $\frac{\sum_{n=1}^{N}\tilde{L}_n + \sum_{p=1}^{P}\tilde{R}_{p,h}}{N+P}$ is equal to the average load of BS_h's partners after receiving traffic. This thesis defines $\tilde{L}_{pars} = \frac{\sum_{n=1}^{N}\tilde{L}_n + \sum_{p=1}^{P}\tilde{R}_{p,h}}{N+P}$.

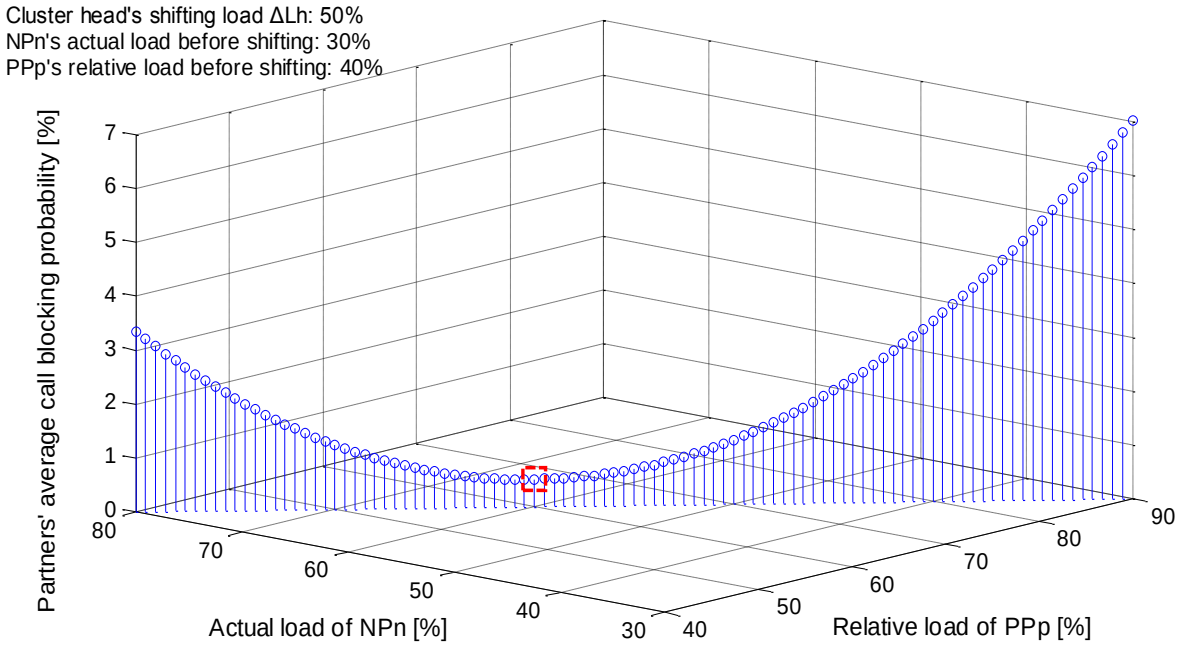
4) Solution to minimise partners' average call blocking probability

According to the theoretical analysis from Equation (4.13) to Equation (4.22), Equation (4.22) is the solution to the optimisation objective of minimising partners' average call blocking probability, which is presented in Equations (4.13)-(4.16). Equation (4.22) means that *each public partner's relative load and non-public partner's actual load reach the same load*. Namely, $\tilde{L}_1 = ... = \tilde{L}_n = ... = \tilde{R}_{p,h} = ... = \tilde{R}_{P,h} = \tilde{L}_{pars}$.

From the analysis above, the partners' average call blocking probability is minimised when the cluster head shifts its traffic until each public partner's relative load and each non-public partner's actual load become equal. This is illustrated in Figure 4.12(a).



(a) Illustration of the solution to minimise partners' average call blocking probability



(b) Example of partners' minimal call blocking probability

Figure 4.12 Illustration of traffic offloading optimisation

Furthermore, under the cluster head BS_h's shifting load ΔL_h, its partners' theoretical minimal call blocking probability $\tilde{B}_{pars}^{\min}$ is

$$\tilde{B}_{pars} = \frac{\sum_{n=1}^{N} \tilde{L}_n \tilde{B}_n + \sum_{p=1}^{P} \tilde{R}_{p,h} \tilde{B}_{p,h}}{\tilde{L}_{pars}^{all}} \quad \xrightarrow{\tilde{L}_n, \tilde{R}_{p,h} = \frac{\Delta L_h + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h}}{N+P}} \quad \tilde{B}_{pars}^{\min} = \frac{\left[\frac{\Delta L_h + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h}}{N+P} M \right]^M \Big/ M}{\sum_{k=0}^{M} \left[\frac{\Delta L_h + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h}}{N+P} M \right]^k \Big/ k!} \tag{4.23}$$

Figure 4.12(b) exemplifies the theoretical minimal call blocking probability. Considering the scenario, where the cluster head BS_h's shifting load $\triangle L_h$=50%, spectrum bandwidth is 5MHz with 25PRB [KAK10]. NP_n's actual load before shifting is 30% and PP_p's relative load before shifting is 40%. As shown in Figure 4.12(b), partners' average call blocking probability $\tilde{B}_{pars}$ reaches the theoretical minimal call blocking probability $\tilde{B}_{pars}^{\min}$, when both PP_p's relative load $\tilde{R}_{p,h}$ and NP_n's actual load $\tilde{L}_n$ are equal to $\tilde{L}_{pars}$ ($\tilde{L}_{pars}$=60%, $\tilde{B}_{pars}^{\min}$=0.501087%, see the red square).

4.5.2.3 Intra-Cluster Shifting Traffic Calculation

Based on the solution to minimise partners' average call blocking probability, the shifting traffic calculation formula is designed in this section.

After receiving the traffic of ΔM_h ($\Delta M_h = \Delta L_h \times M$), the average load of BS_h's partners $\tilde{L}_{pars}$ is

$$\tilde{L}_{pars} = \frac{\sum_{n=1}^{N} \tilde{L}_n + \sum_{p=1}^{P} \tilde{R}_{p,h}}{N+P} = \frac{\frac{\Delta M_h}{M} + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h}}{N+P} = \frac{\Delta L_h + \sum_{p=1}^{N} L_p + \sum_{p=1}^{P} R_{p,h}}{N+P} \tag{4.24}$$

where L_n is NP_n's actual load before traffic shifting, $R_{p,h}$ is PP_p's relative load towards BS_h before traffic shifting.

1) Shifting traffic to PP_p

As discussed in Section 4.5.1.1, $M_{p,h}^{LB}$ is PP_p's load balancing subcarriers for BS_h. BS_h estimates $M_{p,h}^{LB}$ based on the relative load $R_{p,h}$. Equation (4.9) shows that PP_p allocates $R_{p,h} \times M$ subcarriers to both PP_p's serving users, and cluster head BS_j's shifting traffic ($j \in \{1,2 \ldots H\}$, $j \neq h$). Meanwhile, PP_p's actual load cannot exceed the heavily loaded threshold L_{HL}. Hence, BS_h estimates $M_{p,h}^{LB}$ as

$$M_{p,h}^{LB} \approx L_{HL} \times M - R_{p,h} \times M \tag{4.25}$$

The amount of shifting traffic from BS_h to PP_p is defined as $\Delta M_{h,p}$. In order to avoid PP_p being heavily loaded, $\Delta M_{h,p}$ cannot exceed $M_{p,h}^{LB}$, and this constraint is designed as (4.27).

Based on the solution to minimise partners' average call blocking probability, PP_p should receive BS_h's traffic until its relative load $R_{p,h}$ reaches $\tilde{L}_{pars}$. This requirement is designed as (4.26). Hence, BS_h uses (4.26) and (4.27) to calculate $\Delta M_{h,p}$.

$$\Delta M_{h,p} = (\tilde{L}_{pars} - R_{p,h}) \times M \qquad p \in \{1...P\} \tag{4.26}$$

$$\text{Subject to} \quad \Delta M_{h,p} \le M_{p,h}^{LB} \approx (L_{HL} - R_{p,h}) \times M \tag{4.27}$$

2) Shifting traffic to NP_n

In order to reach $\tilde{L}_{pars}$, the amount of shifting traffic from BS_h to NP_n, $\Delta M_{h,n}$ is calculated using (4.28) and (4.29). The constraint (4.29) keeps that $\Delta M_{h,n}$ is less than NP_n's receiving traffic threshold M_n^{LBthr}, in order to avoid the non-public partner being heavily loaded ($M_n^{LBthr} = (L_{HL} - L_n) \times M$. M_n^{LBthr} has similar idea as PP_n's receiving traffic threshold, as discussed in Equation (4.6)).

$$\Delta M_{h,n} = (\tilde{L}_{pars} - L_n) \times M \qquad n \in \{1...N\} \tag{4.28}$$

$$\text{Subject to} \quad \Delta M_{h,n} \le (L_{HL} - L_n) \times M \tag{4.29}$$

4.5.2.4 Cell-Specific Handover Offset Adjustments

Based on the amount of the required shifting traffic, BS_h offloads relevant users to *Partner s* (*Partner s* can be *public partner* or *non-public partner* in BS_h's cluster), by adjusting the cell-specific $HO_{off}(h,s)$. Then U_k *in* BS_h will be offloaded to *Partner s*, if its $RSRP_{k,h}$ from BS_h and $RSRP_{k,s}$ from *Partner s* meet the handover condition [KAPTK10] [3GPP12]

$$RSRP_{k,s} + HO_{off}(h,s) > RSRP_{k,h} + HO_{hys} \tag{4.30}$$

where HO_{hys} is the handover hysteresis, $HO_{hys}=2dB$ [LGK10].

Due to users random channel condition, BS_h adjusts $HO_{off}(h,s)$ with the step size θ ($HO_{off}(h,s)=HO_{off}(h,s)+\theta$, $\theta=1dB$) to offload users. The $HO_{off}(h,s)$ adjustment will stop under two conditions: when the number of releasing subcarriers of offload users reaches the amount of the required shifting traffic; when $HO_{off}(h,s)$ reaches the maximum handover offset HO_{off}^{max} ($HO_{hys}=2dB$, $HO_{off}^{\max}=9dB$ and $HO_{off}^{\max}-HO_{hys}=7dB$ [LGK10] [SOCRATES10]).

4.5.3 Signalling Load and Complexity

This section analyses the signalling load of cooperative traffic shifting, its process is shown in Figure 4.9. In the inter-cluster cooperation, RLRM requires exchanging the actual load or relative load among cells, and hence RLRM consumes extra signalling load. In the intra-cluster cooperation, each cluster head calculates the amount of shifting traffic on the basis of the actual load/relative load, which was obtained in the inter-cluster cooperation. Meanwhile, a cluster head estimates public partner's load balancing subcarriers based on the relative load without extra information exchanges.

In the inter-cluster cooperation, the complexity of RLRM is $H\times O(H^2)$ to calculate PP_p's relative load/s towards H different cluster heads. In the intra-cluster cooperation, the complexity of calculating the average load of BS_h's partners is $O(N+P)$. The complexity of calculating the amount of shifting traffic/s of P public partners and N non-public partners is $2\times P\times O(1)+2\times N\times O(1)$. Hence, the complexity of the intra-cluster cooperation is $O(N+P)+(2N+2P)\times O(1)$.

4.6 Performance Analysis

The proposed self-organising cluster-based cooperative load balancing scheme is evaluated by the system-level simulation platform designed in Chapter 3. The key parameters of this simulator are introduced in Section 3.7.1. The simulator generates 3 hot-spot areas, including 13 hot-spot cells, as shown in Figure 4.13.

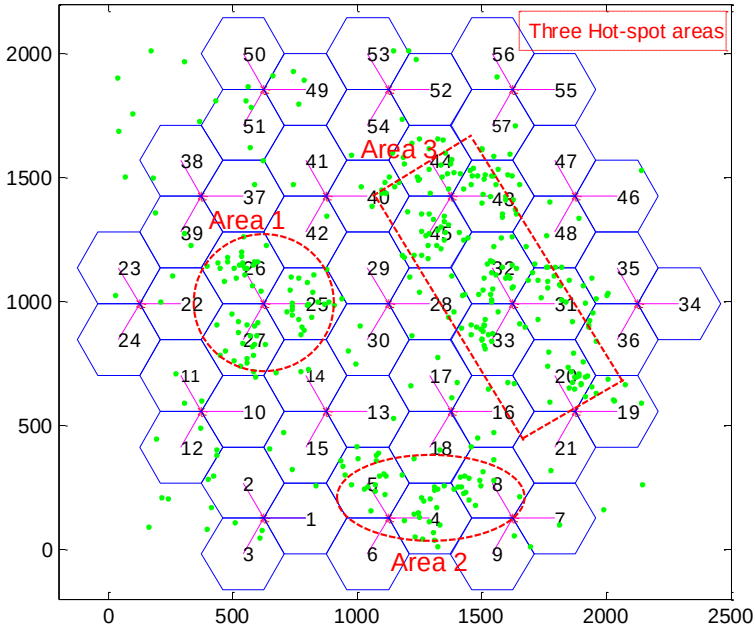


Figure 4.13 Simulation scenario for CCLB (unit: *meter*)

As introduced above, the proposed CCLB scheme includes:

- User-vote assisted clustering algorithm (partner selection stage);
- Cooperative traffic shifting algorithm (traffic shifting stage).

The two algorithms will be evaluated in Section 4.6.1 and Section 4.6.2, respectively.

4.6.1 User-Vote Assisted Clustering

The proposed user-vote assisted clustering algorithm is evaluated in this section. This section tries to clarify the networks performance improvement is due to effective partner selection instead of effective traffic shifting. Therefore, in the user-vote assisted clustering algorithm and the load based clustering algorithm, their traffic shifting stages refer to (4.31) to adjust HO_{off} between cluster head and each partner in the load balancing cluster. Then, cluster head's edge users will shift to partners.

For *public partner* p $p \in \{1...P\}$: $$HO_{Off}(h,p) = f(L_h - L_p) \Rightarrow (L_h - L_p) \times HO_{off}^{max} \quad (4.31a)$$

For *non-public partner* n $n \in \{1...N\}$: $$HO_{Off}(h,n) = f(L_h - L_n) \Rightarrow (L_h - L_n) \times HO_{off}^{max} \quad (4.31b)$$

where L_p and L_n are the actual load of PP_p and NP_n, respectively, L_h is the actual load of the cluster head BS_h, HO_{off}^{max} is the maximum handover offset.

First, Figure 4.14 evaluates the performance of the proposed user-vote assisted clustering algorithm in dealing with the virtual partner problem. The maximum number of partners in each cluster is set to *one*. The load based clustering algorithm, which selects partner based on the neighbouring cell's load, is simulated for comparison. Specifically, in load based clustering algorithm, the cluster head selects one lowest load neighbouring cell as partner, and then the cluster head adjusts its HO_{off} with this partner based on their actual load difference, as shown in Equation (4.31a) and (4.31b).

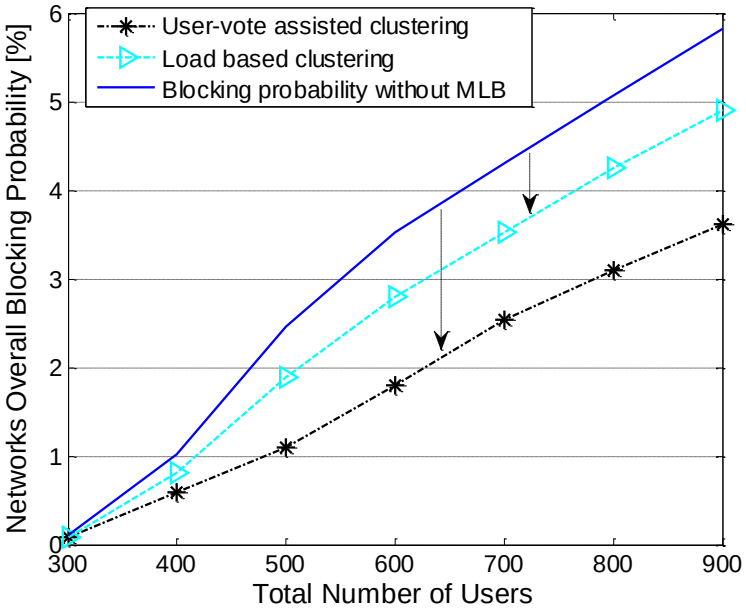


Figure 4.14 Overall call blocking probability Vs Number of users

Call blocking probability is a widely used load balancing performance indicator [NA07] [TY08] [KAPTK10] [SOCRATES10], since the new call users can easily achieve access to the networks under balanced load distribution. As shown in Figure 4.14, the overall call blocking probability increases with adding the total number of users in the networks. However, the proposed user-vote assisted clustering algorithm can lead to lower call blocking probability than load based clustering algorithm. The proposed algorithm further reduces call blocking probability by nearly 1%, compared with the load based clustering algorithm. Therefore, the user-vote assisted clustering algorithm outperforms the conventional load based clustering algorithm, because it can effectively address the virtual partner problem.

Second, Figure 4.15 and Figure 4.16 compare the proposed clustering algorithm's performance under different cluster sizes, namely the maximum number of partners in each cluster sets {0, 1, 2, 3, 4, 5, 6}. The aim is to demonstrate that the proposed algorithm can precisely rank neighbouring cell's load balancing capability to receive cluster head's load, and the proposed algorithm can achieve good load balancing performance with a small number of partners.

Figure 4.15 shows the total number of shifted users under different cluster sizes. The cluster head can shift a large number of users to this partner when choosing one largest selection priority neighbouring cell as the partner. In addition, the cluster head can further shift many users if each cluster head selects two largest priority neighbouring cells as partners. However, the traffic shifting capability improvement is limited under three or more partners.

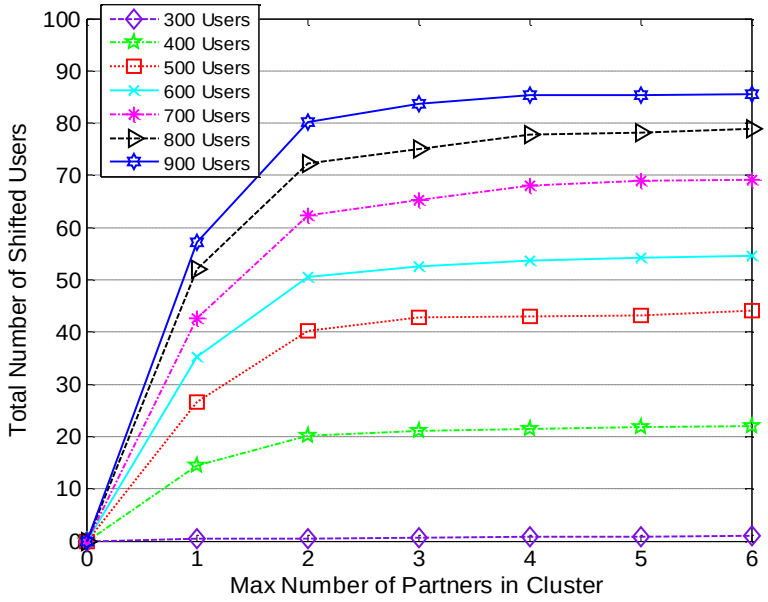


Figure 4.15 Effect of cluster size on total number of shifted users

Corresponding to Figure 4.15, Figure 4.16 depicts that the proposed algorithm can efficiently reduce the overall call blocking probability when each cluster head chooses the highest priority neighbouring cell as the partner. The blocking probability can be further reduced if more high priority neighbouring cells are chosen as partners, but the further reduction is limited under three/four/five/six partners. For example, under 900users scenario, one-partner cluster can reduce the blocking probability from 5.9% to 3.6%, and two-partner cluster can further reduce it to 2.95%, while three/four/five/six-partner cluster can slightly reduce the blocking probability until 2.85% of six-partner cluster.

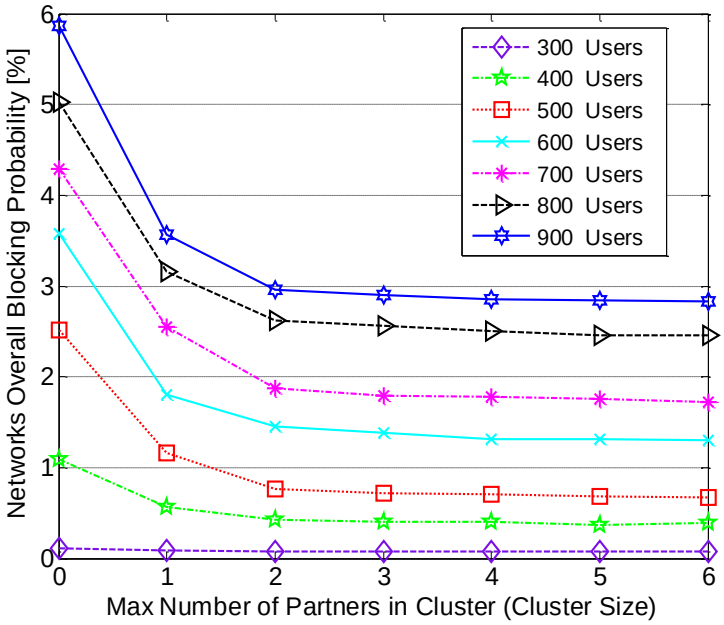


Figure 4.16 Effect of cluster size on overall blocking probability

Figure 4.15 and Figure 4.16 show that partner's priority calculated in Equation (4.3)-(4.5) can precisely reflect neighbouring cell's load balancing capability to receive cluster head's traffic. For example, the highest priority neighbouring cell has the best load balancing capability and the second highest priority neighbouring cell has medium load balancing capability.

The traffic shifting stage requires HO_{off} adjustments. Figure 4.17 compares the number of HO_{off} adjustments in the user-vote two-partner cluster and that in the typical MLB scheme of [NA07]. (In [NA07], a hot-spot cell selects all lightly loaded neighbouring cells as partners and adjusts HO_{off}.) The vertical axis is $\frac{\textit{Number of } HO_{off} \textit{ adjustments in user-vote two-partner cluster}}{\textit{Number of } HO_{off} \textit{ adjustments in typical MLB[NA07]}}$. The number of HO_{off} adjustments in our proposed user-vote two-partner cluster is much less than the MLB scheme of [NA07]. For example, under scenarios with 700 to 900 users, the two-partner cluster only requires 40% HO_{off} adjustments of that in [NA07], which means that our scheme can reduce nearly 60% HO_{off} adjustments.

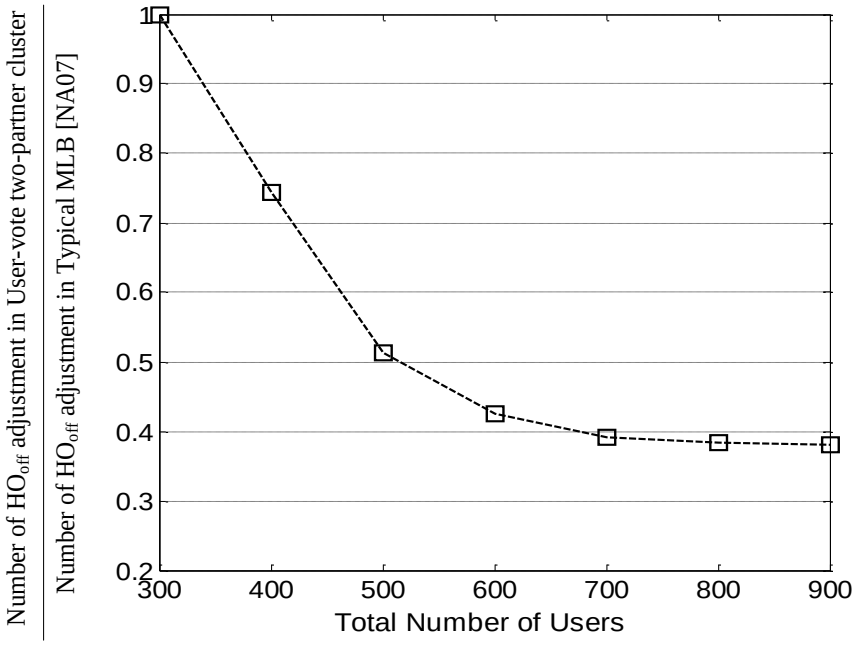


Figure 4.17 Number of HO_{off} adjustments comparison

From Figure 4.15 – Figure 4.17, the proposed algorithm shows that choosing the two best partners can reach a similar load balancing performance as choosing three or more partners. In addition, a two-partner cluster can reduce the unnecessary HO_{off} adjustments. Based on this, we can conclude that the appropriate cluster size is to have one cluster head with *two partners*.

In summary, Figure 4.14 – Figure 4.17 show that the proposed clustering algorithm can deal with the virtual partner problem. They also show the proposed scheme can select a small number of partners (e.g., two partners) to reach a good performance.

4.6.2 Cooperative Traffic Shifting

Since Figure 4.15 and Figure 4.16 show that the two high priority partners can reach a good performance, Table 4.1 shows that *13* cluster heads employ user-vote assisted clustering algorithm to select their two best neighbouring cells as partners. Then, the networks have *8* public partners denoted by *. These public partners and their assisting cluster heads are shown in Table 4.2.

Table 4.1 Cluster structure in single-hop networks simulation

Cluster head	Partner	*Cluster head*	Partner
*Cell*4	Cell9 *Cell18	*Cell*27	*Cell14 *Cell22
*Cell*5	Cell1 *Cell18	*Cell*31	Cell35 *Cell48
*Cell*8	Cell21 *Cell18	*Cell*32	*Cell28 *Cell48
*Cell*20	Cell19 Cell36	*Cell*33	*Cell28 Cell17
*Cell*25	*Cell14 *Cell42	*Cell*43	*Cell48 *Cell57
*Cell*26	*Cell22 *Cell42	*Cell*44	*Cell40 *Cell57
* *Public partner*		*Cell*45	*Cell40 *Cell28

Table 4.2 Public partner's assisting cluster head in single-hop networks simulation

Public partner	Assisting *Cluster head*	Public partner	Assisting *Cluster head*
*Cell18	*Cell*4, *Cell*5, *Cell*8	*Cell28	*Cell*32, *Cell*33,*Cell*45
*Cell14	*Cell*25, *Cell*27	*Cell40	*Cell*44, *Cell*45
*Cell22	*Cell*26, *Cell*27	*Cell48	*Cell*31, *Cell*32,*Cell*43
*Cell42	*Cell*25, *Cell*26	*Cell57	*Cell*43, *Cell*44

This section evaluates the *cooperative traffic shifting*, including its two key mechanisms:

a) Inter-cluster cooperation: RLRM;

b) Intra-cluster cooperation: traffic offloading optimisation.

In order to evaluate the proposed RLRM in addressing the aggravating load problem, the actual-load based MLB scheme is simulated for comparison, under the same clusters structure of Table 4.1.

Figure 4.18 shows the average load of public partners after traffic shifting. The actual-load based MLB scheme results in many heavily loaded public partners, e.g., under scenarios with 600 to 900 users. While in the cluster-based cooperative load balancing (CCLB) scheme,

the average load of public partners is lower than the heavily loaded threshold L_{HL}. This is because the relative load coordinates multiple clusters' traffic shifting requests.

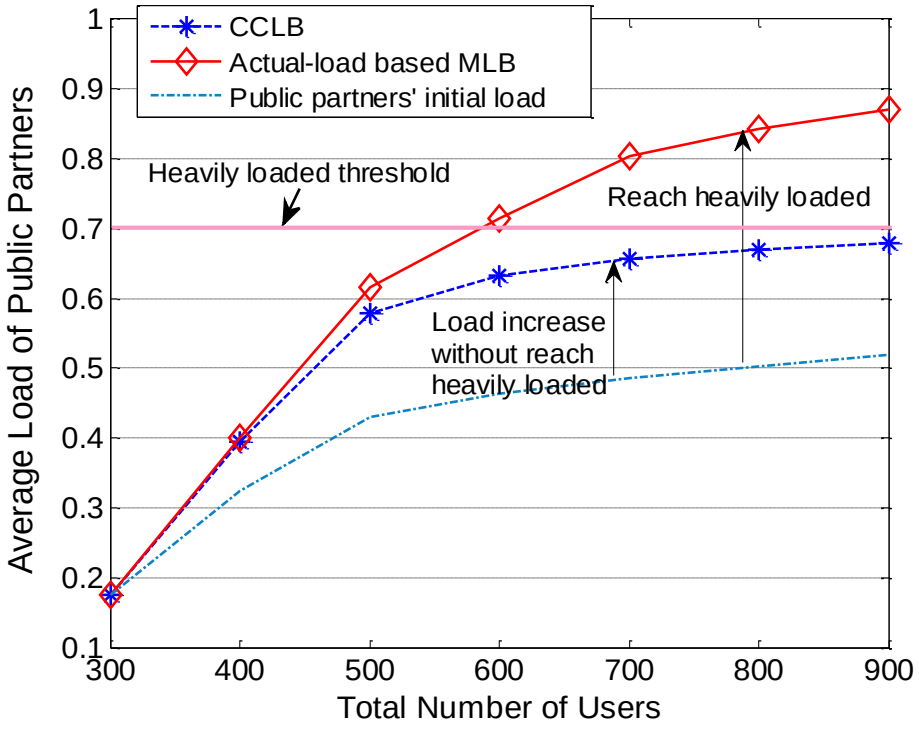


Figure 4.18 Public partners' average load comparison

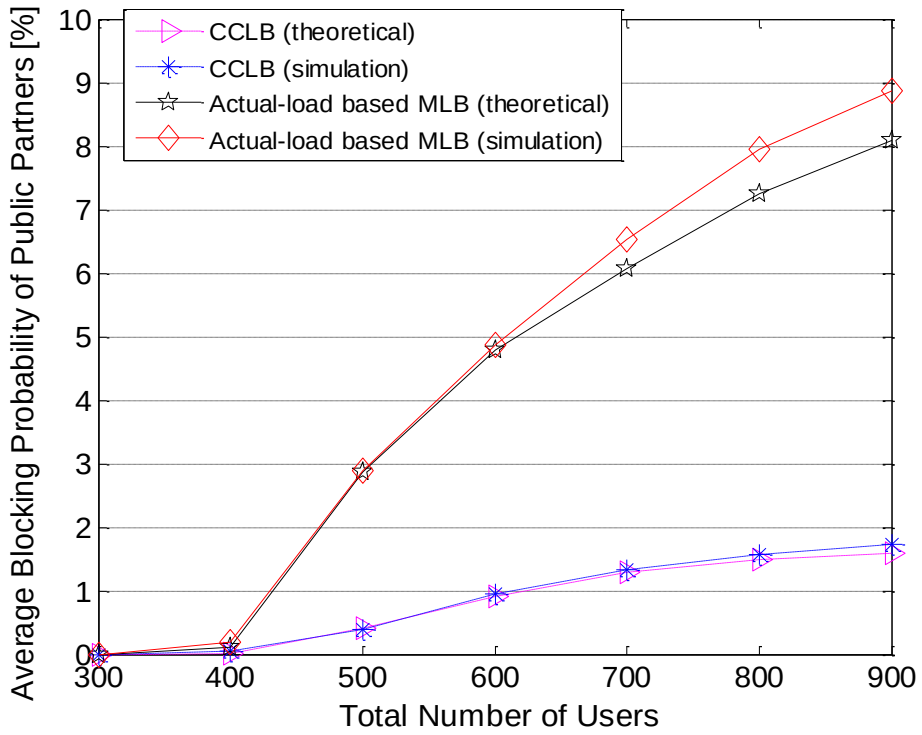


Figure 4.19 Public partners' average blocking probability comparison

Figure 4.19 shows both the simulation and numerical results of all public partners' average call blocking probability after traffic shifting. The numerical results are based on the shifting traffic calculation formula in Equation (4.26) (4.27). Since the proposed CCLB scheme can mitigate public partners becoming heavily loaded, they have sufficient subcarriers to serve new call users. As a result, after traffic shifting, the average call blocking probability is lower than 2%, which is much less than the actual-load based MLB scheme. Figure 4.18 and Figure 4.19 illustrate that the proposed scheme can address the aggravating load problem and keep the public partner's performance at an acceptable level.

In order to evaluate RLRM performance in using the public partner's idle spectrum to reduce cluster heads' load, Figure 4.20 shows the average load of cluster heads after traffic shifting. The *BS state analysis based MLB* scheme [ZQMZ10a] is simulated for comparison, in which a lightly loaded cell can share only one cluster head's load at a time. Hence, this reference scheme can avoid the appearance of a public partner. The CCLB scheme has a better capability to reduce cluster heads' load than the *BS state analysis based MLB* scheme. For example, our scheme can further reduce nearly 10% load, under scenarios with 500 to 900 users. This is because that RLRM allows the appearance of public partner and RLRM can effectively allocate the public partner's idle spectrum to serve each cluster head's shifting traffic.

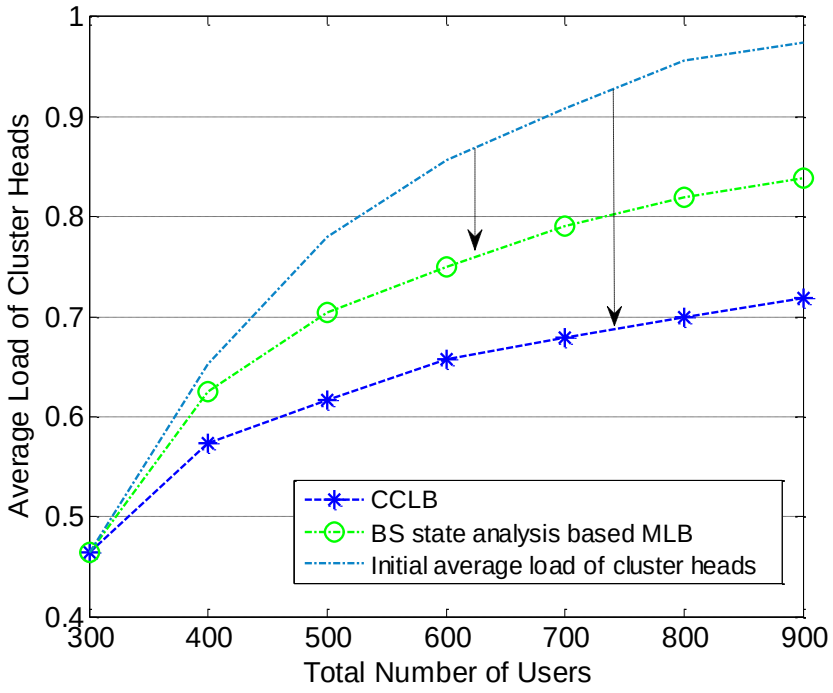


Figure 4.20 Average load of cluster heads comparison

Finally, Figure 4.21 depicts partners' average call blocking probability in each cluster, under 900 users in networks. It compares two algorithms' performance, including the proposed traffic offloading optimisation algorithm and the load difference based traffic shifting algorithm. The load difference based traffic shifting algorithm scheme is discussed in Section 4.9. Figure 4.21 shows that the proposed algorithm achieves much lower call blocking probability, with respect to the load difference based traffic shifting algorithm.

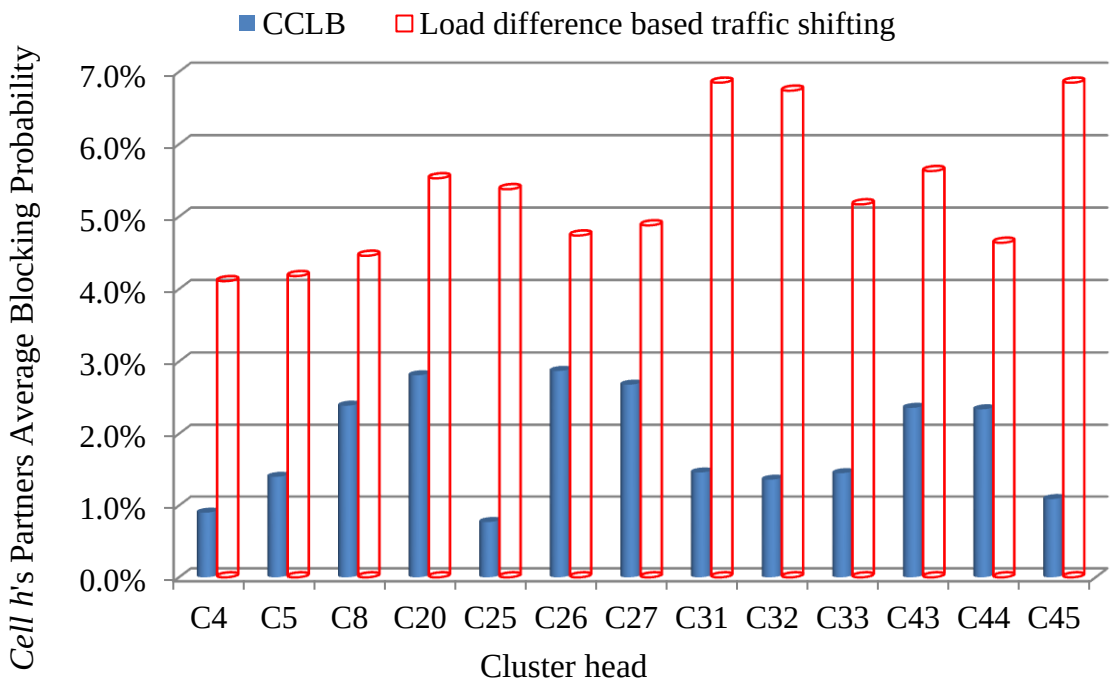


Figure 4.21 Partners' average call blocking probability in each cluster

4.7 Summary

In this chapter, the self-organising cluster-based cooperative load balancing (CCLB) scheme is proposed for single–hop OFDMA cellular networks.

In the load balancing clustering stage, the hot-spot cell employs the user-vote assisted clustering algorithm, which considers the users' channel condition, to select suitable partner cells. Simulation results show that the user-vote assisted clustering algorithm can address the virtual partner problem. Meanwhile, this algorithm can select two best partner cells to effectively balance the load, and reduce the number of HO_{off} adjustments by nearly 60%.

In the traffic shifting stage, a cooperative traffic shifting algorithm is designed. This algorithm consists of the inter-cluster cooperation and the intra-cluster cooperation. In the inter-cluster cooperation, the public partner employs the relative load response model, in order to coordinate multiple clusters' traffic shifting requests and address the aggravating load problem. In the intra-cluster cooperation, the traffic offloading optimisation algorithm minimises the partners' average call blocking probability in each cluster. Simulation results show that the proposed scheme can keep the public partner's load lower than the heavily loaded threshold. The proposed scheme also has much lower partners' average call blocking probability than the load difference based traffic shifting scheme.

4.8 Appendix: The Analysis of $\eta=4$ in User-Vote Model

***The analysis of $\eta=4$* (in User-vote model of Equation (4.2))**

In Equation (4.2), U_k tries to set an appropriate $SINR_{k,h}/\eta$ to identify cell edge user, and to calculate U_k's *vote* towards neighbouring BS_i. Hence, the *3dB* cell edge user identification criterion of [FSCK10] [SKMNT10] is used, as

$$(RSRP_{k,h})_{dB} - (RSRP_{k,i})_{dB} \le 3dB \quad \Rightarrow \quad \frac{(RSRP_{k,h})_{linear}}{(RSRP_{k,i})_{linear}} \le 2 \tag{4.32}$$

where $RSRP_{k,h}$ is from the cluster head BS_h, and $RSRP_{k,i}$ is from neighbouring BS_i. *3dB* denotes their RSRP ratio is 2 times in linear format. Then, we analyse its SINR relationship. The RSRP and SINR in (4.33) and (4.34) are in the linear format.

$$SINR_{k,h} \ge \frac{RSRP_{k,h}}{RSRP_{k,i} + \sum_{\bar{i}=1,\bar{i}\neq i}^{I} RSRP_{k,\bar{i}}} = \frac{2 \times RSRP_{k,i}}{0.5 \times RSRP_{k,h} + \sum_{\bar{i}=1,\bar{i}\neq i}^{I} RSRP_{k,\bar{i}}} \tag{4.33}$$

$$\approx 4 \times \frac{RSRP_{k,i}}{RSRP_{k,h} + \sum_{\bar{i}=1,\bar{i}\neq i}^{I} RSRP_{k,\bar{i}}} = 4 \times SINR_{k,i}^{est} \tag{4.34}$$

where $SINR_{k,h}$ is U_k's serving SINR from BS_h, $SINR_{k,i}^{est}$ is U_k's SINR estimation towards BS_i.

(4.33) sets '$\geq$', since $RSRP_{k,i} + \sum_{\bar{i}=1,\bar{i}\neq i}^{I} RSRP_{k,\bar{i}}$ is the theoretical heaviest overall interference of $SINR_{k,h}$. In (4.34), '$\approx$' denotes approximately, since if U_k is shifted, $RSRP_{k,h}$ from the cluster head becomes the heaviest interference, compared with $RSRP_{k,\bar{i}}$ from other neighbouring BSs. Therefore, $\eta=4$ is a suitable value in Equation (4.2) to calculate vote.

4.9 Appendix: Load Difference based Traffic Shifting

Load difference based traffic shifting (in Simulation of Figure 4.21)

As introduced in Section 4.5.2, the cluster head BS_h tries to release ΔM_h subcarriers, which is flexible according to different load balancing objectives. The simulator of CCLB assumes BS_h's objective L_h^* is equal to the average load of its cluster, namely $L_h^* = (L_h + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h})/(1+N+P)$. Therefore, BS_h's load reduction ΔL_h equals $L_h - L_h^*$. BS_h's releasing subcarriers ΔM_h can be expressed as

$$\Delta M_h = M \times \Delta L_h = M \times (L_h - \frac{L_h + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h}}{1+N+P}) \tag{4.35}$$

Under the condition of releasing ΔM_h subcarriers, the traffic offloading optimisation algorithm refers to Section 4.5.2.3 to calculate the amount of shifting traffic to each partner.

Since the traffic shifting direction is from a hot-spot cell to each partner, many MLB schemes [NA07] [KAPTK10] adjust HO_{off} between the hot-spot cell and each partner, based on their load difference. However, the simulation comparison cannot directly use the equations in these load difference schemes [NA07] [KAPTK10]. It is because their equations do not pre-define an overall load reduction ΔL_h of the cluster head under two or more partners, and the more partners will lead to larger load reduction of the cluster head. For example, Equation (4.31), which is based on the general principle of HO_{off} adjustment in [NA07], has no constraint of cluster head's overall load reduction.

The simulator tries to avoid the cluster head having different load reduction objectives, in the conventional 'load difference' scheme and our 'cluster-based cooperative load balancing' scheme. Hence, we follows the basic idea of load difference and designs the 'load difference

based traffic shifting' scheme, in which the cluster head always has a certain overall load reduction ΔL_h under different numbers of partners. In this scheme, for a particular *partner*, the amount of shifting traffic $\Delta M_{h,p}$ or $\Delta M_{h,n}$ is based on the actual load between BS_h and this *Partner*:

Amount of shifting traffic from BS_h to PP_p: $$\Delta M_{h,p} = \frac{(L_h - L_p)}{\sum_{\check{p}=1}^{P}(L_h - L_{\check{p}}) + \sum_{\check{n}=1}^{N}(L_h - L_{\check{n}})} \times \Delta M_h$$

Amount of shifting traffic from BS_h to NP_n: $$\Delta M_{h,n} = \frac{(L_h - L_n)}{\sum_{\check{p}=1}^{P}(L_h - L_{\check{p}}) + \sum_{\check{n}=1}^{N}(L_h - L_{\check{n}})} \times \Delta M_h$$

where L_h is the actual load of the cluster head, L_p and L_n are the actual load of PP_p ($p \in \{1..P\}$) and NP_n ($n \in \{1..N\}$), respectively. ΔM_h is BS_h's total releasing subcarriers calculated in (4.35). Then, based on the amount of the required shifting traffic, the cluster head adjusts cell-specific handover offset to offload users towards each public partner/non-public partner. In addition, the two schemes in Figure 4.21 have the same cluster structure as shown in Table 4.1.

Chapter 5 Load Balancing in Multi-Hop Cellular Networks

5.1 Introduction

In future LTE-Advanced networks, fixed relay is an important technology to extend the cell coverage and enhance users' performance in cell edge area [3GPP10e]. The deployment of RSs increases the complexity and brings challenges, since most shifted users are served by partner's RSs. In Section 5.2, the features and challenges of load balancing in fixed relay cellular networks are investigated. Then, the CCLB scheme is modified to apply in fixed relay cellular networks. The CCLB scheme aims at effectively shifting traffic and addressing the RS aggravating load problem in fixed relay cellular networks. Simulation results show that the user-vote assisted clustering algorithm can select a small number of partners to effectively balance load. The relay-level user shifting algorithm can mitigate the heavily loaded RS and reduce the load balancing handover rate by nearly 20%.

In the scenarios where fixed relay is not deployed, the user shifted from hot-spot BS may receive weak signal, due to the far distance propagation loss from partner's BS. This will result in the shifted user's link quality degradation. In Section 5.3, a user relaying model, in which a non-active user is treated as relay to forward signal to the shifted user, is employed to address the link quality degradation problem. Based on the user relaying model, user relaying assisted traffic shifting (URTS) scheme is proposed. In URTS scheme, the shifted user can select a suitable non-active user as relay user, in order to effectively enhance shifted user's link quality with low cost of relay user's energy consumption. Simulation results

show that URTS scheme can improve the SINR of shifted users by 20%~75%, and reduce the load balancing handover failure rate. URTS scheme also reaches a good trade-off between shifted user's performance improvement and relay user's rate loss.

5.2 Cluster-Based Cooperative Load balancing in Fixed Relay Cellular Networks

This section investigates the problems and challenges faced by load balancing in fixed relay cellular networks. Besides the virtual partner problem and the aggravating load problem of public partner, load balancing in fixed relay cellular networks confronts another problem: RS aggravating load problem. To deal with these problems, the CCLB scheme is modified to apply in fixed relay cellular networks. The user-vote assisted clustering algorithm and the relative load response model are modified to address the virtual partner problem and the aggravating load problem of public partner. A novel relay-level user shifting algorithm, which analyses RS's spectrum usage and users' channel condition, is proposed to address the RS aggravating load problem.

5.2.1 System Model of Fixed Relay Cellular Networks

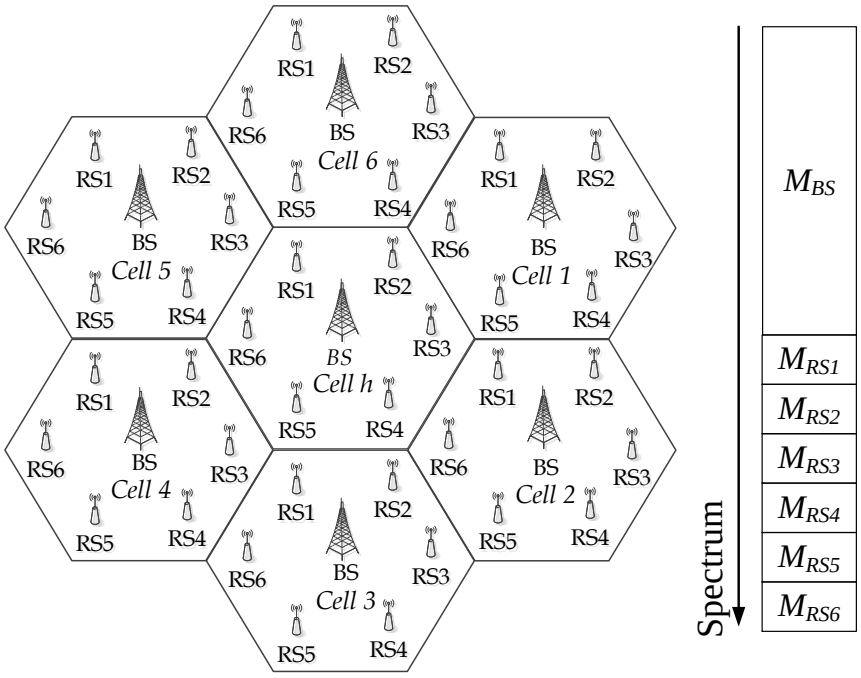


Figure 5.1 Layout and frequency planning in fixed relay cellular networks

The system model of fixed relay cellular networks is described firstly. The system model is the foundation for the problems analysis in Section 5.2.2 and the CCLB scheme designed in Section 5.2.3. Figure 5.1 shows the layout of fixed relay cellular networks. As introduced in Section 3.3.2, RS is located at 2/3 of cell radius [CJC09]. The modified soft frequency reuse (MSFR) technology [GZLLZ07] is employed to pre-allocate spectrum resources to each BS/RS node. The inner users are served by BS with M_{BS} spectrum. The edge users are served by RS via two-hop transmission. The BS-RS link and the RS-edge user link are allocated the same M_{RS} spectrum at different time slots [GZLLZ07].

RS works in decode-and-forward (DF) mode. As introduced in Section 3.5.3, BS and RS are located on the rooftop, and BS-RS link are LOS transmission with good channel condition. Therefore, RS can decode signal successfully, re-encode and transmit signal to users [WTJLHL10] [FW11].

For a more general system model, it is assumed that the hot-spot $Cell_h$ had I neighbouring cells indexed with i ($i \in \{1,2 \ldots I\}$), and $Cell_h$'s RSs serve K users indexed with k ($k \in \{1,2 \ldots K\}$). U_k is served by $RS_{sev,h}$. The definitions and system parameters that will be used in Section 5.2 are listed as follows:

$Cell_h$: Cluster head $Cell_h$. As discussed in Section 4.4.1, BS_h discovers itself as a cluster head if the period, when its actual load is higher than the threshold L_{HL}, is larger than the critical time T_{crit}.

$Cell_i$: Neighbouring $Cell_i$. $Cell_h$ has I neighbouring cells indexed with i ($i \in \{1,2 \ldots I\}$).

$Cell_j$: Partner $Cell_j$. Partners are a subset of neighbouring cells, which are selected by the cluster head to share the load. Partners consist of public partners and non-public partners.

Public partner, non-public partner: Public partner receives traffic from multiple cluster heads. Non-public partner receives traffic from one cluster head. Assuming $Cell_h$ has N non-public partners indexed with n($n \in \{1,2..N\}$), and P public partners indexed with p ($p \in \{1,2..P\}$).

L_{node} : Actual load. To a specific Cell/BS/RS node, the actual load is defined as the ratio of the number of subcarriers in use to the node's total number of pre-allocated subcarriers. $L_{node} = M_{use\ node} / M_{node}$ and $0\% \le L_{node} \le 100\%$.

L_{HL}: The threshold of heavy load. L_{HL} = 70%, as introduced in Section 2.1.2.

$RS_{sev,h}$: The serving RS of U_k. $RS_{sev,h}$ is located in $Cell_h$.

$RS_{r,i}$: RS_r in neighbouring $Cell_i$ $r \in \{1,2 \ldots 6\}$.

$RS_{t,j}$: RS_t in partner $Cell_j$.

$RSRP_{k,RS_{sev,h}}$: U_k's reference signal received power from serving $RS_{sev,h}$.

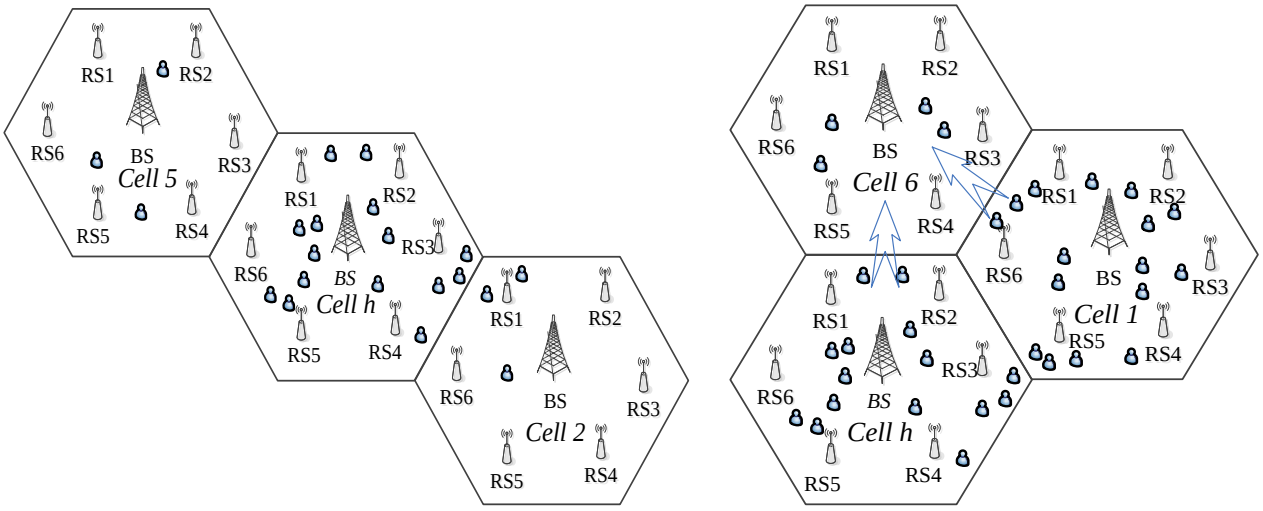
$RSRP_{k,RS_{r,i}}$: U_k's reference signal received power from neighbouring $RS_{r,i}$.

$SINR^{est}_{k,RS_{r,i}}$: U_k's SINR estimation towards $RS_{r,i}$.

5.2.2 Problem Formulation

In Chapter 4, the virtual partner problem and the aggravating load problem of public partner have been discussed in load balancing of single-hop cellular networks. The two problems still exist in load balancing of fixed relay cellular networks. In addition, in fixed relay cellular networks, load balancing suffers a particular problem: RS aggravating load problem.

5.2.2.1 Virtual Partner and Aggravating load of Public Partner



(a) Virtual partner problem (b) Aggravating load problem of public partner

Figure 5.2 Virtual partner problem, aggravating load problem of public partner in fixed relay networks

- Virtual partner problem: As exemplified in Figure 5.2(a), $Cell_h$ is a hot-spot and intends to shift some traffic out. Applying the load based partner selection, both $Cell_5$ and $Cell_2$ appear to be possible partners with the same priorities as they have the same load. However, $Cell_2$ is more suitable. $Cell_5$ is a *virtual partner*, because $Cell_h$'s users are far from

$Cell_5$'s RSs, and few users are able to be shifted to $Cell_5$.

- Aggravating load problem of public partner: In fixed relay cellular networks, multiple hot-spot cells may shift traffic to one lightly loaded cell, which becomes a public partner. As exemplified in Figure 5.2(b), lightly loaded $Cell_6$ is the public partner of hot-spot $Cell_1$ and $Cell_h$. Without the coordination of shifting traffic from $Cell_1$ and $Cell_h$, the public partner $Cell_6$ will become heavily loaded.

In the CCLB scheme, the above two problems are addressed by the modified user-vote assisted clustering algorithm and the relative load response model.

5.2.2.2 Particular Problem: RS Aggravating Load

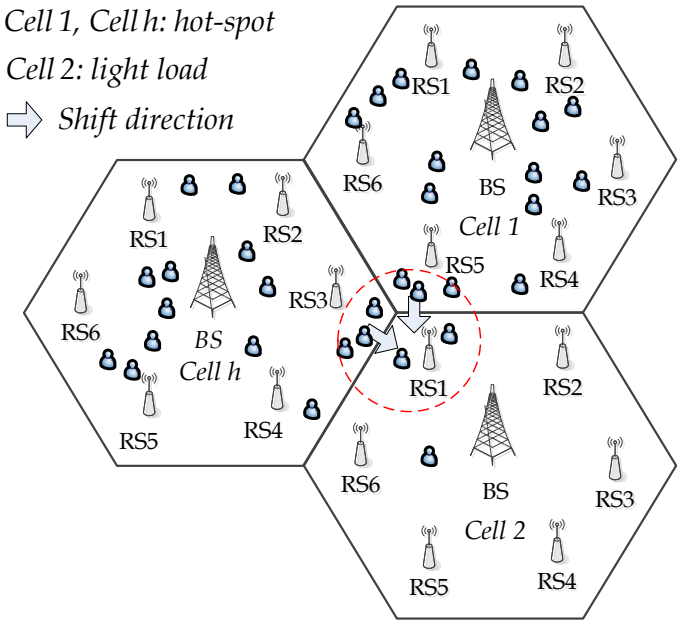


Figure 5.3 RS aggravating load problem

RS aggravating load problem: This is a particular problem in fixed relay cellular networks. When a hot-spot cell shifts edge users to partners, most shifted users will be served by partners' RSs, while each RS is pre-allocated only a small part of spectrum resources in the partner cell. As exemplified in Figure 5.3, $Cell_1$ and $Cell_h$ are hot-spot and try to shift some traffic to the lightly loaded $Cell_2$. Due to the random user distribution, most edge users of $Cell_1$ and $Cell_h$ will be shifted to RS_1 of $Cell_2$. After receiving traffic, RS_1 becomes heavily loaded. The heavily loaded RS is defined as the RS aggravating load problem. This problem affects networks performance because heavily loaded RS_1 may result in the handover failure. Conventional load balancing schemes may suffer the RS aggravating load problem. In the

CCLB scheme, a novel relay-level user shifting algorithm is designed for fixed relay cellular networks. This algorithm jointly considers the spectrum usage of RS and users' channel condition, in order to shift appropriate users from the cluster head to partner's RS. This can deal with the RS aggravating load problem and reduce the handover failure rate.

5.2.3 Process of CCLB Scheme

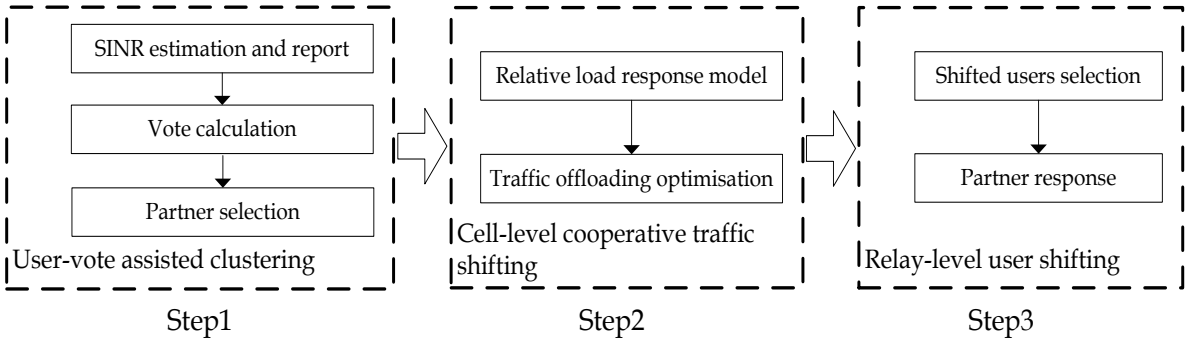


Figure 5.4 Flowchart of CCLB scheme in fixed relay cellular networks

In fixed relay cellular networks, the proposed CCLB scheme follows three steps, as shown in Figure 5.4. Initially, a hot-spot cell identifies itself as cluster head by comparing its actual load and the heavily loaded threshold. Then, the cluster head employs the user-vote assisted clustering algorithm to select suitable neighbouring cells as its partners. Since multiple cluster heads may select one public partner, in the cell-level cooperative traffic shifting stage, the public partner employs the relative load response model to feedback the public partner's relative load to each requesting cluster head. Based on the relative load of the public partner and the actual load of the non-public partner, the cluster head employs traffic offloading optimisation algorithm to calculate the amount of shifting traffic to each partner cell.

In the novel relay-level user shifting stage, the cluster head selects possible shifted users, according to the amount of the required shifting traffic and users' channel condition in target RSs. The cluster head sends these users' information (such as SINR) to the partner. Then the partner cell analyses RSs' idle spectrum and users' channel condition to confirm the shifted users. Finally, the partner and cluster head adjust cell-specific handover offset to offload users.

5.2.4 User-Vote assisted Clustering

In Section 5.2.4, the user-vote assisted clustering is modified to apply in fixed relay cellular networks. The aim is to deal with the virtual partner problem and select appropriate partners to shift traffic in fixed relay cellular networks.

1) SINR Estimation and Report

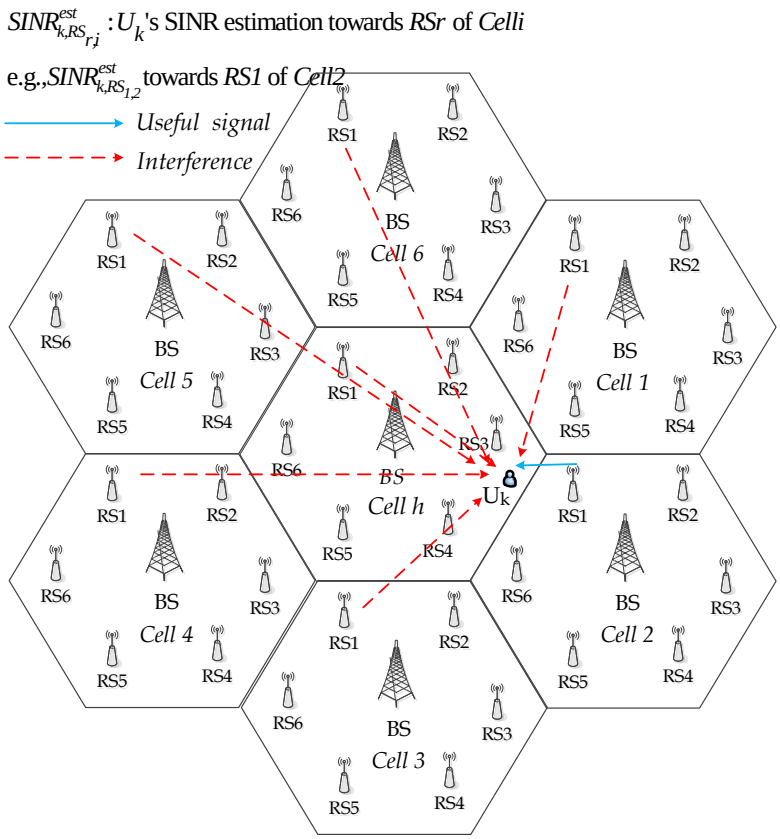


Figure 5.5 Illustration of co-channel interference

From the system model, there are K users served by $Cell_h$'s RSs, and U_k is served by $RS_{sev,h}$. In the user-vote assisted clustering algorithm, U_k estimates its SINR from each RS_r in each neighbouring $Cell_i$ (namely, $RS_{r,i}$ $r \in \{1,2\ldots6\}$ $i \in \{1,2 \ldots I\}$). According to the system model, each cell pre-allocates the co-channel M_{RSr} subcarriers to RS_r. As exemplified in Figure 5.5, subcarriers pre-allocated to RS_1 in $Cell_2$ are reused with RS_1 in all neighbouring cells. In addition, the precise SINR estimation is difficult because U_k's subcarriers allocated by $RS_{r,i}$ is time-varying, based on the channel condition. Therefore, U_k estimates the worst SINR from $RS_{r,i}$ using (5.1). $SINR^{est}_{k,RS_{r,i}}$ reflects the channel condition after U_k is shifted, and is used to

calculate U_k's *vote*.

$$SINR^{est}_{k,RS_{r,i}} = \frac{RSRP_{k,RS_{r,i}}}{\sum_{\substack{\tilde{i}=1 \\ \tilde{i}\neq i}}^{I} RSRP_{k,RS_{r,\tilde{i}}} + RSRP_{k,RS_{r,h}}} \tag{5.1}$$

where $RSRP_{k,RS_{r,i}}$ is from the voting target $RS_{r,i}$, $RSRP_{k,RS_{r,h}}$ is from the corresponding RS_r of the cluster head $Cell_h$, $\sum_{\substack{\tilde{i}=1 \\ \tilde{i}\neq i}}^{I} RSRP_{k,RS_{r,\tilde{i}}}$ is from the corresponding RS_r of other neighbouring cells [3GPP10d]. For the worst SINR, the noise is negligible compared with the interference in the full frequency reuse networks.

* Note that in user-vote assisted clustering algorithm of fixed relay cellular networks, user estimates SINR from neighbouring cell's RSs, because RSs are located in cell edge to be able to serve the shifted user. In user-vote assisted clustering algorithm of single-hop cellular networks, user estimates SINR from neighbouring cell's BS, because shifted user is served by BS directly.

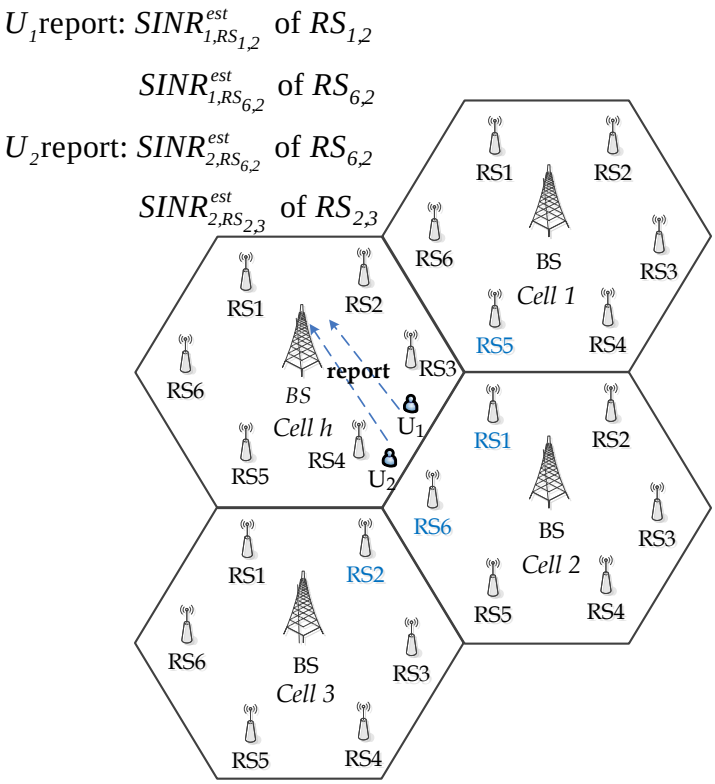


Figure 5.6 Illustration of SINR report (Report two largest SINR neighbouring RSs)

After SINR estimation, U_k only reports estimated *SINR* of the two neighbouring RSs with the

largest $SINR^{est}_{k,RS_{r,i}}$, to the cluster head. This step is illustrated in Figure 5.6. It is due to the fact that in most cases, U_k is near two neighbouring RSs at most, the received *SINR* from other neighbouring RSs are so small that U_k cannot be shifted to. Besides, U_k also reports the $SINR_{k,RS_{sev,h}}$ received from $RS_{sev,h}$ ($RS_{sev,h}$ is the serving RS of U_k, and $RS_{sev,h}$ is located at the cluster head $Cell_h$.).

* Note that user's vote is calculated in the cluster head $Cell_h$, instead of user itself. This step is different from user-vote assisted clustering in single-hop cellular networks. This is because that user's estimated SINR will also be used in relay-level user shifting stage (step3 of CCLB).

2) Vote Calculation

After receiving U_k's report, based on the serving $SINR_{k,RS_{sev,h}}$ and the estimated $SINR^{est}_{k,RS_{r,i}}$, the cluster head calculates U_k's *vote* as $V_{k,RS_{r,i}}$, using (5.2). $V_{k,RS_{r,i}}$ indicates U_k's probability of being offloaded to $RS_{r,i}$, reflecting U_k's satisfaction degree to $RS_{r,i}$.

$$V_{k,RS_{r,i}} = \begin{cases} 1 & SINR^{est}_{k,RS_{r,i}} \geq SINR_{k,RS_{sev,h}} \\ \dfrac{SINR^{est}_{k,RS_{r,i}}}{SINR_{k,RS_{sev,h}}} & SINR^{est}_{k,RS_{r,i}} < SINR_{k,RS_{sev,h}} \end{cases} \quad (5.2)$$

- For the users with $SINR^{est}_{k,RS_{r,i}} \geq SINR_{k,RS_{sev,h}}$, $RS_{r,i}$ can serve them with satisfactory channel condition. Hence, they vote for $RS_{r,i}$ with full vote $V_{k,RS_{r,i}} = 1$.
- For the users with $SINR^{est}_{k,RS_{r,i}} < SINR_{k,RS_{sev,h}}$, $V_{k,RS_{r,i}}$ equals the ratio of $SINR^{est}_{k,RS_{r,i}}$ to $SINR_{k,RS_{sev,h}}$.

3) Partner Selection

Based on each user's *vote*, the cluster head calculates the total votes TV_i of neighbouring $Cell_i$.

$$TV_i = \sum_{k=1}^{K} \underset{r\in\{1...6\}}{Max} V_{k,RS_{r,i}} \quad (5.3)$$

TV_i reflects the traffic shifting capability of neighbouring $Cell_i$, decided by users' channel condition. The higher the value, the more users in the cluster head $Cell_h$ favour to be shifted to neighbouring $Cell_i$.

In order to select appropriate partner cells, the cluster head also considers the actual load of neighbouring $Cell_i$. Actual load reflects the idle subcarriers of neighbouring $Cell_i$ to serve shifted users. The actual load information can be exchanged between neighbouring cells via the X2 interface [3GPP10a] [3GPP11a]. Then, the above two factors are jointly considered and the selection priority of neighbouring $Cell_i$ is defined as

$$\Pr_i = \frac{TV_i}{K} + (1 - L_i) \qquad i \in \{1 \ldots I\} \tag{5.4}$$

where L_i is the actual load of neighbouring $Cell_i$, (1- L_i) can reflect the idle spectrum of $Cell_i$. K is the total number of users in $Cell_h$'s RSs. The denominator K guarantees that the factor of total votes ranges from 0 to 1, which is in the same magnitude as the actual load.

According to (5.4), under the same number of votes, the neighbouring cell with lower load has higher priority to be selected as a partner. Meanwhile, under the same load, the neighbouring cell with higher votes has higher priority.

A load filter is also deployed to avoid selecting a heavily loaded neighbouring cell:

$$\textit{Load filter:} \qquad L_i < L_{HL} \qquad i \in \{1 \ldots I\} \tag{5.5}$$

The cluster head sorts neighbouring cells, which satisfy the Filter (5.5), in the descending order according to their priorities (5.4). Then the cluster head continuously selects the highest priority neighbouring cell as partner in sequence, until the number of selected cells is larger than the required number of partners (the appropriate number of partners is discussed via simulation analysis in Section 5.2.7.1). Then, the cluster head sends a request message to selected neighbouring cells for cluster construction. The selected neighbouring BS will respond with a confirmation message. The clustering process is finished after partners' response.

5.2.5 Cell-Level Cooperative Traffic Shifting

1) *Relative Load Response Model*

After load balancing clustering stage, the public partner employs the RLRM (relative load response model), which is previously designed in Section 4.5.1. RLRM can coordinate multiple cluster heads' traffic shifting requests and mitigate the probability of public partner being heavily loaded.

The system model assumes that the cluster head $Cell_h$ has N non-public partners indexed with n ($n \in \{1,2 \ldots N\}$), and P public partners indexed with p ($p \in \{1,2 \ldots P\}$). Using RLRM, the relative load of public partners and the actual load of non-public partners are as follows:

- $R_{1,h} \ldots R_{p,h} \ldots R_{P,h}$: Relative load of public partners
- $L_1 \ldots L_n \ldots L_N$: Actual load of non-public partners

2) *Traffic Offloading Optimisation*

After above actual load / relative load response, $Cell_h$ employs the (cell-level) traffic offloading optimisation algorithm to calculate $Cell_h$'s amount of shifting traffic to each partner. The (cell-level) traffic offloading optimisation algorithm is proposed in Section 4.5.2. As discussed in Section 4.5.2, the cluster head can minimise its partners' average call blocking probability, when *each public partner's relative load and non-public partner's actual load reach the same load*. The (cell-level) shifting traffic calculation formulas are as follows:

Assuming $Cell_h$ tries to release ΔM_h subcarriers via shifting users to partners. After receiving the traffic of ΔM_h ($\Delta M_h = \Delta L_h \times M$), the average load of $Cell_h$'s partners is

$$\tilde{L}_{pars} = \frac{\frac{\Delta M_h}{M} + \sum_{n=1}^{N} L_n + \sum_{p=1}^{P} R_{p,h}}{N+P} = \frac{\Delta L_h + \sum_{p=1}^{N} L_p + \sum_{p=1}^{P} R_{p,h}}{N+P} \tag{5.6}$$

a) Shifting traffic from $Cell_h$ to PP_p (*Public partner p*)

For PP_p, $Cell_h$ tries to shift traffic $\Delta M_{h,p}$ to PP_p until PP_p's relative load $R_{p,h}$ reaches $\tilde{L}_{pars}$.

In order to avoid PP_p being heavily loaded, $Cell_h$'s amount of shifting traffic $\Delta M_{h,p}$ cannot exceed $M_{p,h}^{LB}$ ($M_{p,h}^{LB}$ is PP_p's load balancing subcarriers for receiving $Cell_h$'s shifting users). As discussed in (4.25), $Cell_h$ estimates $M_{p,h}^{LB}$ as $M_{p,h}^{LB} \approx L_{HL} \times M - R_{p,h} \times M$. Therefore, $Cell_h$ uses (5.7) and (5.8) to calculate the amount of shifting traffic to PP_p.

$$\Delta M_{h,p} = (\tilde{L}_{pars} - R_{p,h}) \times M \qquad p \in \{1...P\} \tag{5.7}$$

$$\text{Subject to} \quad \Delta M_{h,p} \le M_{p,h}^{LB} \approx (L_{HL} - R_{p,h}) \times M \tag{5.8}$$

b) Shifting traffic from $Cell_h$ to NP_n (*Non-public partner n*)

In order to reach $\tilde{L}_{pars}$, the amount of shifting traffic from $Cell_h$ to NP_n, $\Delta M_{h,n}$ is calculated via (5.9) and (5.10). The constraint (5.10) guarantees that $\Delta M_{h,n}$ is less than NP_n's receiving traffic threshold, in order to avoid NP_n being heavily loaded.

$$\Delta M_{h,n} = (\tilde{L}_{pars} - L_n) \times M \qquad n \in \{1...N\} \tag{5.9}$$

$$\text{Subject to} \quad \Delta M_{h,n} \le (L_{HL} - L_n) \times M \tag{5.10}$$

5.2.6 Relay-Level User Shifting

According to the amount of the required shifting traffic, the cluster head tries to offload edge users to partner cells. However, due to the random user location and RS's limited spectrum, this process may result in the RS aggravating load problem as discussed in Section 5.2.2.2.

In CCLB scheme, a novel relay-level user shifting algorithm is designed. Its aims are:

- Shift appropriate users, which have good channel condition in partners' RSs;
- Mitigate the RS aggravating load problem.

As shown in Figure 5.7, this algorithm includes two steps. From the cluster head side, the cluster head selects part of edge users as possible shifted users, in order to satisfy the

amount of the required shifting traffic to each partner. Then the cluster head reports these selected users' information to the partner. From the partner cell side, the partner analyses its RS's spectrum usage and then confirms handover users. Finally, the partner and the cluster head adjust cell-specific handover offset to offload users.

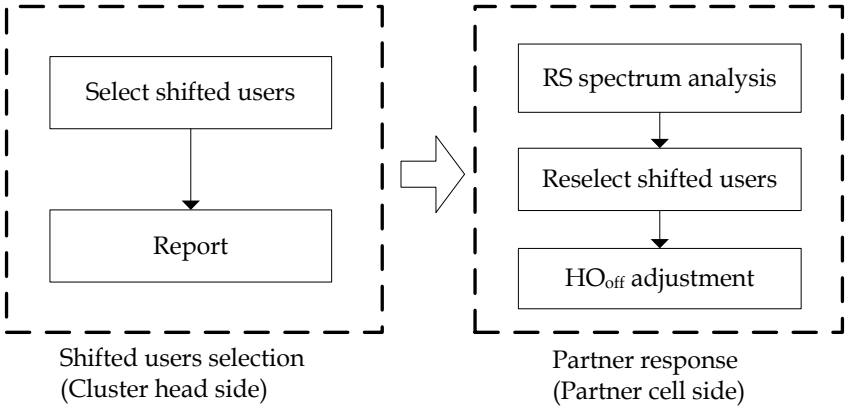


Figure 5.7 Flowchart of relay-level user shifting algorithm

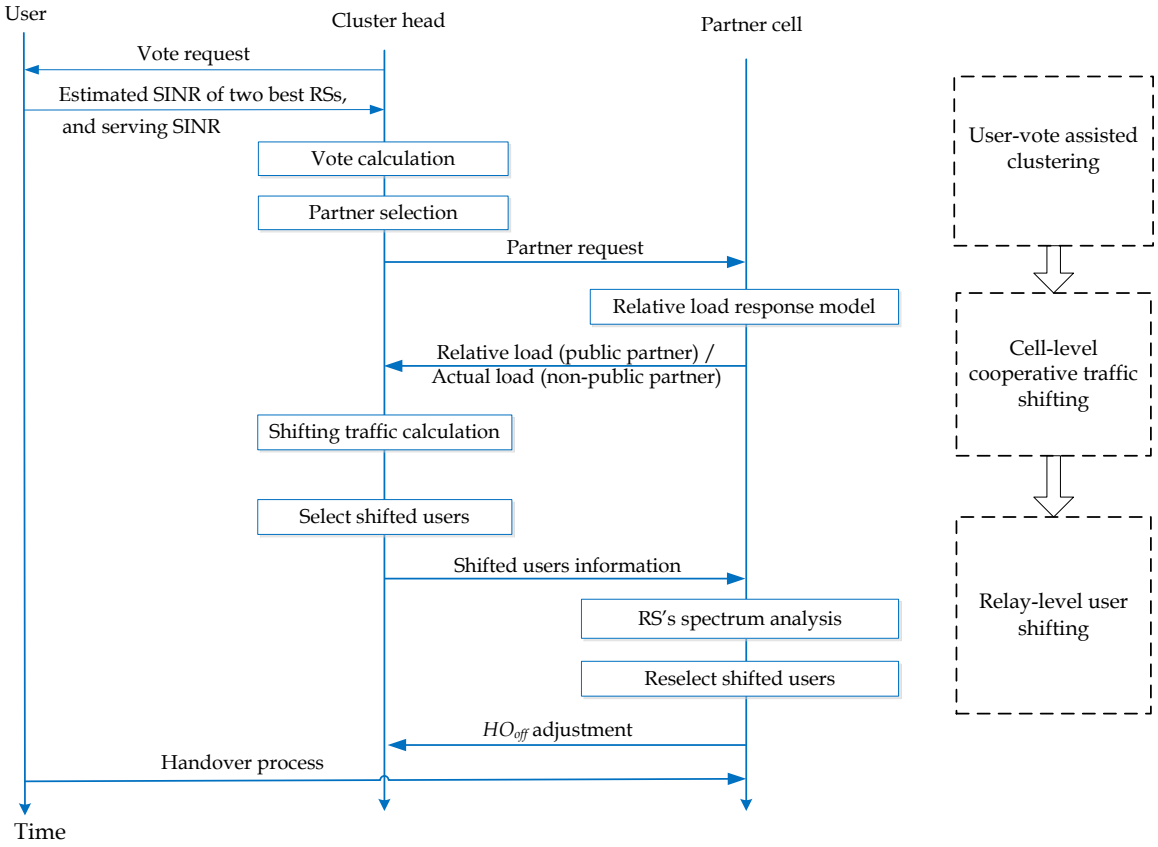


Figure 5.8 Overall process of CCLB scheme in fixed relay cellular networks

1) *Shifted Users Selection*

Figure 5.8 shows the overall process of the CCLB scheme in fixed relay cellular networks. In the user-vote assisted clustering stage, each edge user of the cluster head reports the estimated SINR of target RS. In the relay-level user shifting stage, the cluster head selects shifted users, based on user's estimated SINR of target RS.

Considering the scenario that the cluster head $Cell_h$ selects shifted users to partner $Cell_j$ ($Cell_j$ can be public partner or non-public partner). The amount of the required shifting traffic from $Cell_h$ to partner $Cell_j$ is $\Delta M_{h,j}$. U_k estimates that $RS_{t,j}$ can provide the highest SINR in partner $Cell_j$. Hence, U_k considers $RS_{t,j}$ as the handover target RS. The shifted users selection process is as follows:

All U_k/s are sorted in the descending order, according to their $SINR^{est}_{k,RS_{t,j}}$/s towards target $RS_{t,j}$/s. Then the cluster head $Cell_h$ iteratively selects the user with the highest $SINR^{est}_{k,RS_{t,j}}$, until the amount of selected users' releasing subcarriers satisfies $\Delta M_{h,j}$.

Finally, $Cell_h$ informs partner $Cell_j$ about the selected users' information, including their target $RS_{t,j}$/s, $SINR^{est}_{k,RS_{t,j}}$/s and each user's average number of allocated subcarriers.

2) *Partner Response*

From the partner $Cell_j$ side, $Cell_j$ tries to effectively balance load and avoid heavily loaded $RS_{t,j}$. To achieve the objective, $Cell_j$ analyses $RS_{t,j}$'s spectrum usage, using (5.11)-(5.13).

Assuming $RS_{t,j}$'s receiving traffic is $M^{LB}_{RS_{t,j}}$. (5.11) shows that after receiving traffic, $RS_{t,j}$'s load cannot reach the heavily loaded threshold L_{HL}. Therefore, $RS_{t,j}$'s receiving traffic $M^{LB}_{RS_{t,j}}$ should be less than $(L_{HL} - L_{RS_{t,j}}) \times M_{RS_{t,j}}$, as shown in (5.12). Then, $Cell_j$ calculates $RS_{t,j}$'s receiving traffic threshold $M^{LBthr}_{RS_{t,j}}$ by (5.13).

$$L_{RS_{t,j}} + \frac{M^{LB}_{RS_{t,j}}}{M_{RS_{t,j}}} \le L_{HL} \tag{5.11}$$

$$\Rightarrow \quad M^{LB}_{RS_{t,j}} \le (L_{HL} - L_{RS_{t,j}}) \times M_{RS_{t,j}} \tag{5.12}$$

$$\Rightarrow \quad M_{RS_{t,j}}^{LBthr} = (L_{HL} - L_{RS_{t,j}}) \times M_{RS_{t,j}} \tag{5.13}$$

where $L_{RS_{t,j}}$ is the actual load of $RS_{t,j}$, $M_{RS_{t,j}}$ is the total number of subcarriers in $RS_{t,j}$.

From (5.11)-(5.13), $RS_{t,j}$'s idle subcarriers, which are allocated to shifted users, cannot exceed $M_{RS_{t,j}}^{LBthr}$. Otherwise, $RS_{t,j}$ will become heavily loaded and suffer the aggravating load problem.

After analysing $RS_{t,j}$'s receiving traffic threshold $M_{RS_{t,j}}^{LBthr}$, partner $Cell_j$ reselects the user with the highest $SINR_{k,RS_{t,j}}^{est}$ in sequence, until the amount of reselected users' required subcarriers reaches $M_{RS_{t,j}}^{LBthr}$.

* Note that if $Cell_j$ is a public partner, all requesting users, which are selected by all cluster heads, are sorted in descending order according to user's SINR estimation $SINR_{k,RS_{t,j}}^{est}$. Then $Cell_j$ reselects the user with high $SINR_{k,RS_{t,j}}^{est}$ in sequence, until the amount of reselected users' required subcarriers reaches $M_{RS_{t,j}}^{LBthr}$. This reselection step aims at ensuring $RS_{t,j}$ can serve users of good channel condition and avoiding heavily loaded $RS_{t,j}$, in both the non-public partner scenario and the public partner scenario.

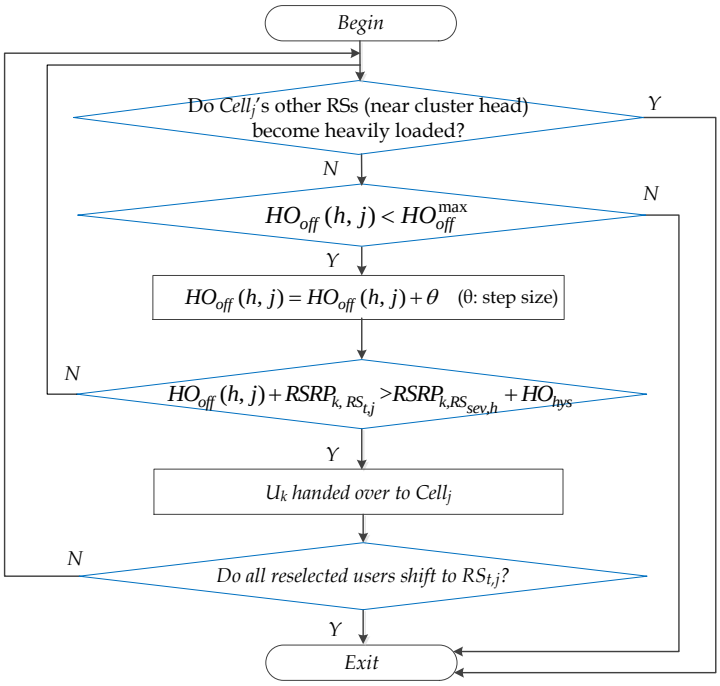


Figure 5.9 HO_{off} adjustments flowchart in relay-level user shifting algorithm

Then, $Cell_j$ adjusts $HO_{off}(h,j)$ between $Cell_h$ and $Cell_j$ to offload the reselected users. The flowchart of HO_{off} adjustment is depicted in Figure 5.9. U_k's RSRP from serving $RS_{sev,h}$ and target $RS_{t,j}$ are denoted as $RSRP_{k,RS_{sev,h}}$ and $RSRP_{k,RS_{t,j}}$, respectively. In order to precisely offload reselected users, $Cell_j$ adjusts $HO_{off}(h,j)$ with the step size θ (θ=1dB), until all reselected users are handed over or $HO_{off}(h,j)$ reaches the maximum handover offset $HO_{off}^{\max}$. $Cell_j$ will also stop $HO_{off}(h,j)$ adjustment, if $Cell_j$'s other RSs receive $Cell_h$'s users and become heavily loaded.

5.2.7 Performance Analysis

The proposed CCLB scheme is evaluated by the system-level simulation platform designed in Chapter 3. The key parameters are introduced in Section 3.7.3. This simulator generates two hot-spot areas, including six heavily loaded cells, as shown in Figure 5.10.

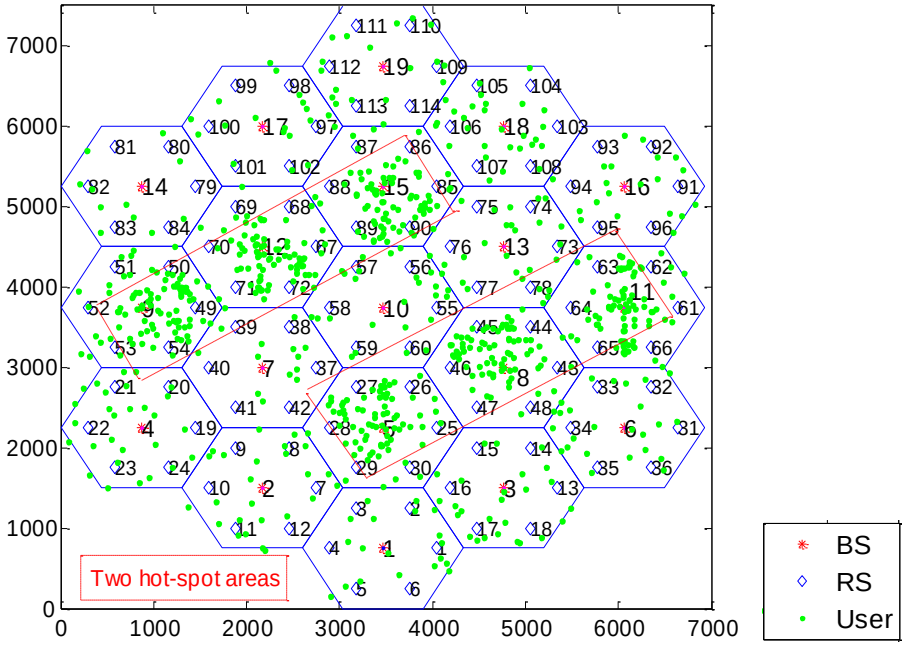


Figure 5.10 Simulation scenario for CCLB scheme in fixed relay cellular networks (unit: *meter*)

The proposed CCLB scheme includes:

- User-vote assisted clustering algorithm (for partner selection);
- Relative load response model (for cell-level cooperative traffic shifting);
- Relay-level user shifting algorithm.

These algorithms will be evaluated in Section 5.2.7.1, Section 5.2.7.2 and Section 5.2.7.3, respectively.

5.2.7.1 User-Vote assisted Clustering

The user-vote assisted clustering algorithm is evaluated in this section. Meanwhile, in its traffic shifting stage, this section refers to the traffic shifting stage in [WTJLHL10] to calculate the amount of shifting traffic and adjust HO_{off}. Then the edge users in the cluster head are handed over to partners. [WTJLHL10] is introduced in Section 2.5.1.

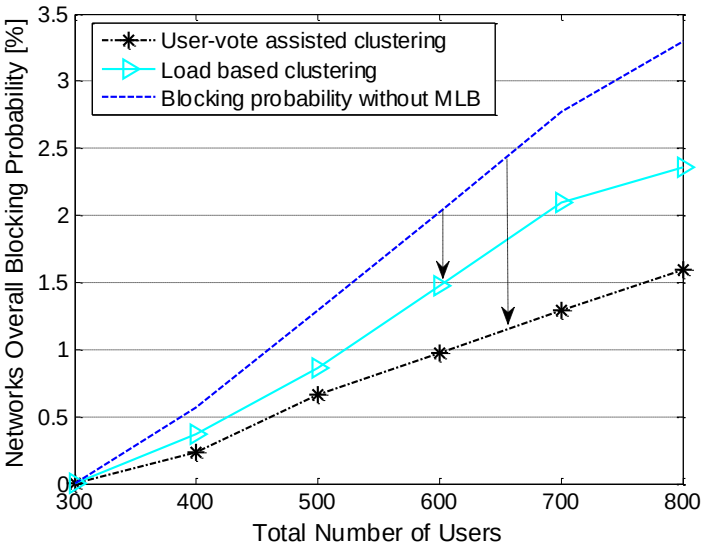


Figure 5.11 Overall call blocking probability comparison in clustering stage (one partner per cluster)

Figure 5.11 evaluates the performance of user-vote assisted clustering algorithm in dealing with the virtual partner problem. The maximum number of partners in each cluster is being set to *one*. The load based clustering algorithm, which selects partner based on the neighbouring cell's load, is simulated for comparison. Specifically, in load based clustering algorithm, the cluster head selects one lowest load neighbouring cell as partner. Figure 5.11 shows that the user-vote assisted clustering algorithm can lead to lower call blocking probability than load based clustering algorithm. For example, under 800 users scenario, the load based clustering algorithm can reduce the call blocking probability by nearly 0.9%, while the user-vote assisted clustering can reduce the call blocking probability by nearly 1.7%. Therefore, the user-vote assisted clustering algorithm outperforms the conventional load based clustering algorithm, because it can effectively deal with the virtual partner problem in fixed relay cellular networks.

In order to demonstrate that user-vote assisted clustering algorithm can achieve good load balancing performance with a small number of partners, we examine its performance under different cluster sizes, namely the maximum number of partners sets {0, 1, 2, 3, 4, 5, 6}.

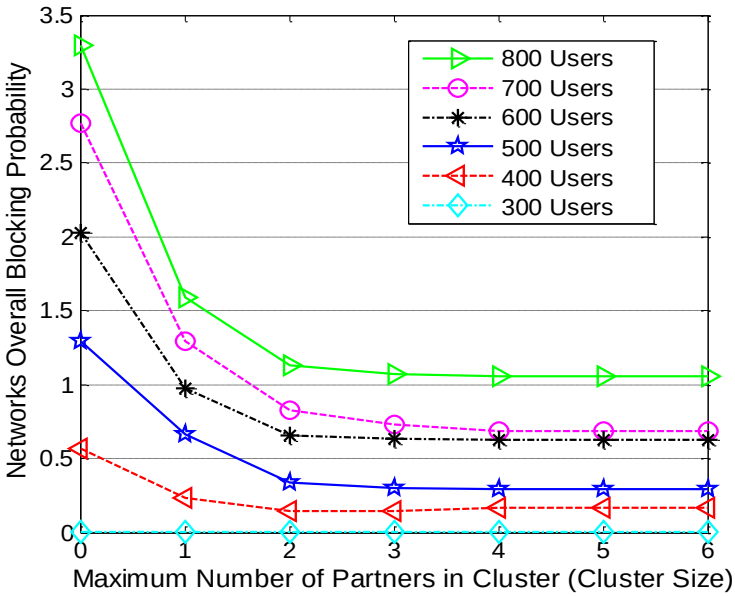


Figure 5.12 Overall call blocking probability in different cluster sizes

Figure 5.12 shows that the proposed clustering algorithm can efficiently reduce the overall call blocking probability when each cluster head chooses the highest priority neighbouring cell as the partner. The call blocking probability can be further reduced if more high priority neighbouring cells are chosen as partners, but the reduction is no obvious when the number of partners in each cluster goes beyond two. For example, under 800 users scenario, one-partner cluster can reduce the blocking probability from 3.3% to 1.6%, and two-partner cluster can further reduce it to 1.13%, while three/four/five/six-partner cluster can slightly reduce the blocking probability until 1.05% of six-partner cluster. Therefore, Figure 5.12 demonstrates that the priority formula (5.4) can precisely reflect neighbouring cell's load balancing capability to receive the cluster head's traffic. Therefore, in this simulation scenario, the appropriate cluster size is two partners.

In summary, Figure 5.11 and Figure 5.12 demonstrate that the proposed user-vote assisted clustering algorithm can deal with the virtual partner problem in fixed relay cellular networks. They also show that the proposed algorithm can effectively rank neighbouring cell's capability to receive cluster head's traffic, and a small number of partners (two partners) can obtain good load balancing performance and improve the clustering efficiency.

5.2.7.2 Relative Load Response Model

In Section 5.2.7.1, it is demonstrated that the two high priority partners can reach good load balancing performance. Table 5.1 shows that *6* cluster heads employ the user-vote assisted clustering algorithm to select their *two* appropriate neighbouring cells as partners. Then, there are *3* public partners denoted by *.

Table 5.1 Cluster structure in fixed relay cellular networks simulation

Cluster head	Partner	*Cluster head*	Partner
*Cell*5	*Cell7 *Cell10	*Cell*11	Cell6 Cell16
*Cell*8	*Cell10 *Cell13	*Cell*12	*Cell7 *Cell10
*Cell*9	Cell4 *Cell7	*Cell*15	*Cell10 *Cell13
* *Public partner*			

In order to evaluate RLRM performance in mitigating the aggravating load problem of public partner, the actual load based scheme is simulated under the same clusters structure of Table 5.1.

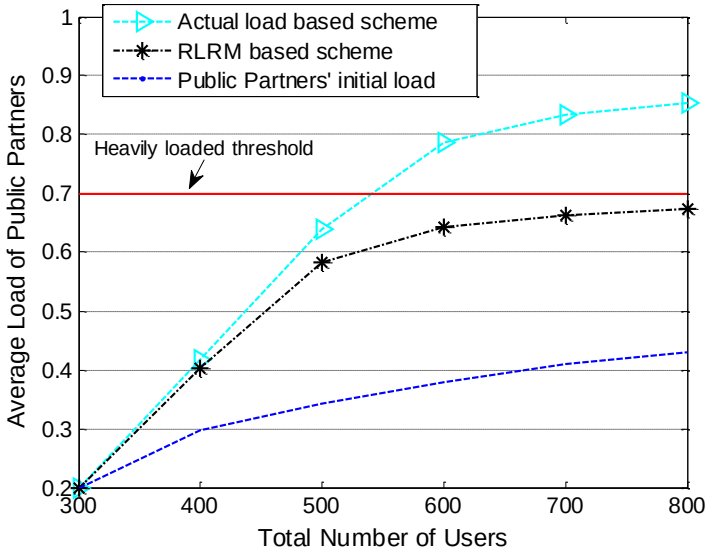


Figure 5.13 Public partners' average load comparison in fixed relay networks

Figure 5.13 shows the average load of public partners after traffic shifting. The actual load based scheme results in many heavily loaded public partners, e.g., under scenarios with 600-800 users. While using the RLRM based scheme, the average load of public partners is lower

than the heavily loaded threshold L_{HL}. This is because that RLRM coordinates multiple clusters' traffic shifting requests and the public partner's idle spectrum. Therefore, RLRM works not only in single-hop cellular networks but also in fixed relay cellular networks.

5.2.7.3 Relay-Level User Shifting Algorithm

In order to evaluate the performance of relay-level user shifting algorithm, both CCLB scheme and CCLB without relay-level user shifting scheme are simulated. The two schemes have the same cluster structure in Table 5.1, and the two schemes employ cell-level cooperative traffic shifting stage. Besides, the traffic load balancing scheme of [WTJLHL10] is also simulated for comparison.

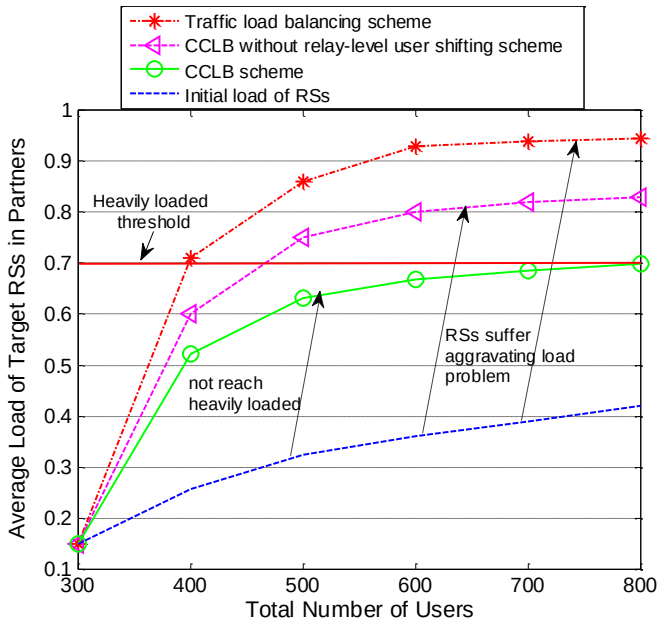


Figure 5.14 Comparison of average load of target RSs[2]

Figure 5.14 shows the average load of partners' target RSs in different schemes. The reference traffic load balancing scheme [WTJLHL10] results in the RS aggravating load problem. In CCLB without relay-level user shifting scheme, the average load of RSs is lower than that in traffic load balancing scheme, because RLRM can reduce the amount of shifting traffic towards the public partner. However, CCLB without relay-level user shifting scheme also might result in the RS aggravating load problem.

Figure 5.14 shows that the proposed CCLB scheme keeps RSs' average load lower than the

[2] Target RSs: The RSs, which are in partner cells and serve the shifted users.

heavily loaded threshold. This is because that the relay-level user shifting algorithm considers both RS's spectrum usage and users' channel condition, thus addressing RS aggravating load problem.

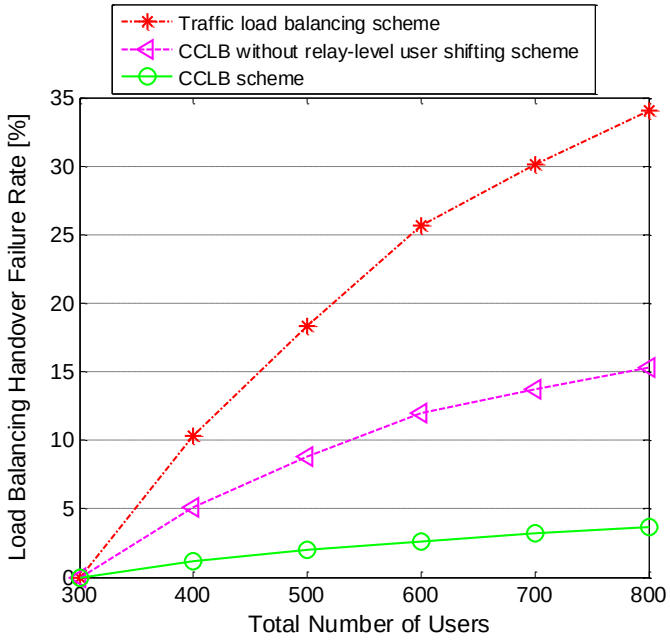


Figure 5.15 Load balancing handover failure rate comparison in fixed relay networks

Figure 5.15 shows the load balancing handover failure rate in different schemes. In fixed relay cellular networks with MSFR technology [GZLLZ07], a shifted user may suffer load balancing handover failure when the target RS in the partner cell cannot provide sufficient subcarriers to meet the shifted user's service requirement, or the partner cell cannot provide the shifted user with the required SINR to sustain connection [JBTMK10] [SOCRATES10]. The traffic load balancing scheme [WTJLHL10] has the highest handover failure rate, since a heavily loaded RS does not have enough subcarriers to satisfy the shifted user's service requirement. The CCLB without relay-level user shifting scheme has medium handover failure rate. The handover failure rate is significantly reduced by the proposed CCLB scheme, since the relay-level user shifting algorithm mitigates heavily loaded RS. Hence, RS has sufficient subcarriers to satisfy the shifted user's service requirement. For example, under scenarios with 600 to 800 users, the handover failure rate in CCLB scheme is nearly 20% lower than that of traffic load balancing scheme.

Overall, from the simulation analysis above, the proposed CCLB scheme can deal with the virtual partner problem via user-vote assisted clustering algorithm, and mitigate the public partner's aggravating load problem via RLRM. Furthermore, it can mitigate RS aggravating

load problem and reduce load balancing handover failure rate via relay-level user shifting algorithm. In summary, the proposed CCLB scheme can be applied in both single-hop cellular networks and fixed relay cellular networks.

5.3 User Relaying assisted Traffic Shifting Scheme

In cellular networks without fixed relay deployment, such as single-hop cellular networks, the shifted user's received signal power from the partner cell may be lower than that from the hot-spot cell (cluster head). Therefore, the shifted user may suffer the link quality degradation. After CCLB scheme implementation, we employ a user relaying model and propose user relaying assisted traffic shifting (URTS) scheme to deal with the link quality degradation. In URTS scheme, the shifted user selects a suitable non-active user as relay user to forward signal, in order to obtain the diversity gain and enhance the shifted user's link quality. Since the user relaying model consumes relay user's energy, a utility function is designed in the relay selection stage of URTS scheme, in order to improve the shifted user's link quality with low cost of relay user's energy consumption.

5.3.1 Problem Formulation

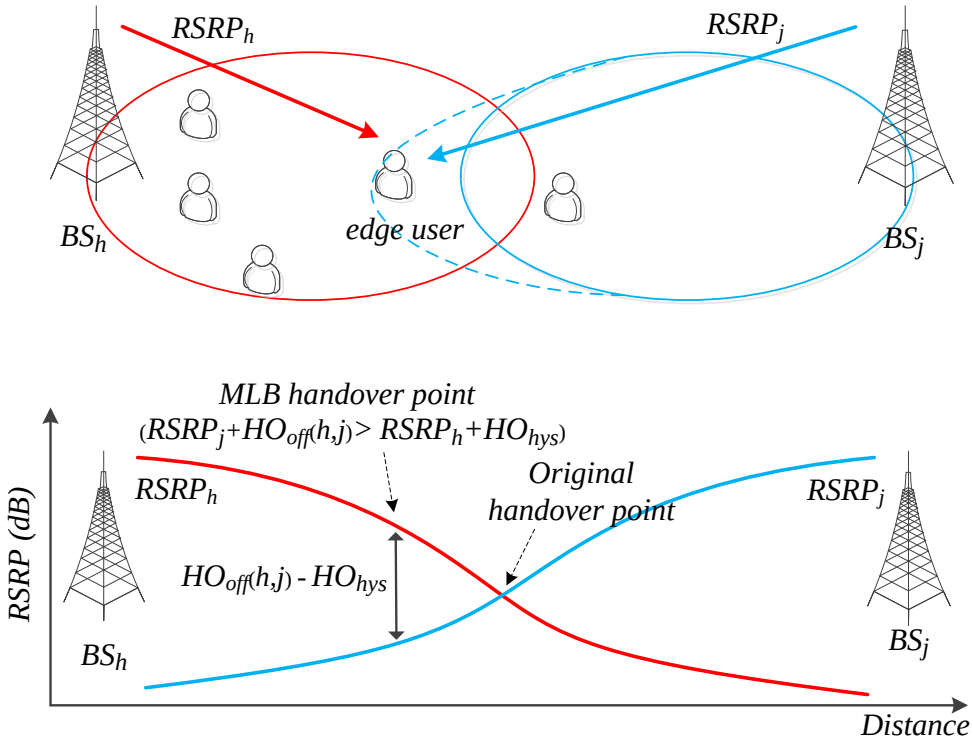


Figure 5.16 Illustration of handover condition in MLB

MLB can shift hot-spot cell's users to neighbouring cells. However, shifted users may receive low RSRP and suffer link quality degradation. As shown in Figure 5.16, BS_h is a hot-spot and

tries to shift users to the lightly loaded BS_j. Then BS_h increases HO_{off} value towards BS_j. Finally, the edge user in BS_h can satisfy the hard handover condition $RSRP_j + HO_{off}(h,j)_{dB} \geq RSRP_h + HO_{hys}$ and shift to BS_j. Figure 5.16 shows that the shifted user receives lower $RSRP_j$ after MLB, compared with $RSRP_h$ before MLB. Furthermore, the reduced RSRP may result in low SINR.

In this thesis, the phenomenon of the reduced RSRP and even reduced SINR of shifted users is called *link quality degradation*. This problem impacts networks performance. The shifted user may experience handover failure due to poor link quality. In addition, after successful handover, BS_j needs to assign more subcarriers to meet the shifted user's requirement.

In order to deal with the link quality degradation and improve networks performance, a user relaying model is employed: a non-active user is employed as a mobile relay to forward data to a shifted user. The spatially independent transmission path (relay link, BS direct link) can achieve diversity gain to enhance the shifted user's link quality.

5.3.2 User Relaying Model

When a user is in idle mode [3GPP10c] [3GPP11d], it is called the non-active user in this thesis. In the downlink of each non-active user, the control channel is partial used while the traffic channel is idle. Hence, a non-active user can forward signal to a shifted user.

Assuming a shifted user is originally served by a hot-spot cell. In MLB, the user is shifted to a lightly loaded cell. Then, the basic idea of user relaying model is depicted in Figure 5.17. In the user relaying model, the shifted user selects a non-active user located in the lightly loaded cell as the relay user. When BS transmits data to the shifted user, the relay user also receives the data at the first time slot (TS) and then forwards to the shifted user at the second time slot.

In order to simplify the description, it is assumed that the system model includes a shifted user, defined as *User u*; several non-active users, defined as *Relay r* $r\epsilon\{1 \ldots R\}$; and a lightly loaded BS, defined as *BS b*. Therefore, after relay selection, a specific user relaying model consists of one *Relay r*, one shifted *user u* and one source *BS b*.

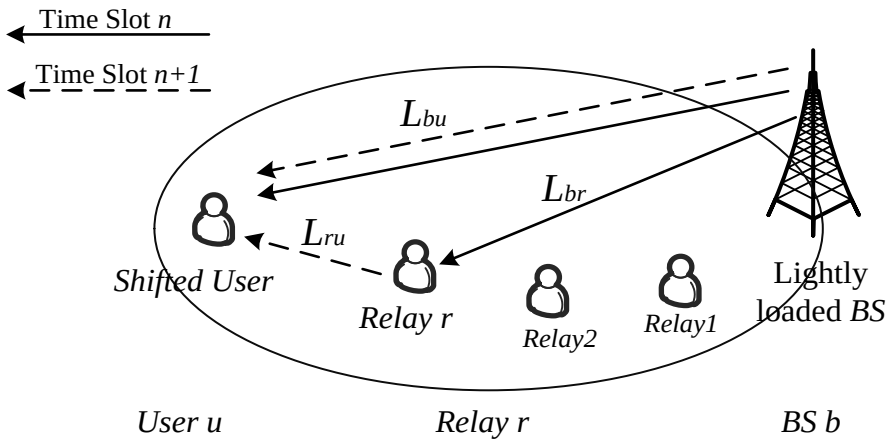


Figure 5.17 User relaying model

The downlink transmission mode is shown in Figure 5.17, including two consecutive time slots [CYOCY08] [WJL09]. At TS *n*, both *User u* and *Relay r* listen to the transmission of *BS b*. At TS *n+1*, both BS *b* and Relay *r* transmit signal to *User u* simultaneously. Note that *BS b* transmits the identical data at two consecutive time slots [CYOCY08] [WJL09] [WTJLHL10] in the user relaying model.

In addition, *Relay r* operates in the amplify-and-forward (AF) mode [JXJA08]. In the AF mode, the relay user amplifies all received signals, including interference, noise and user signal. Then it forwards these signals to the shifted user. The AF mode suits the user device, since the AF mode is more simple to implement, and requires lower computation capability than the DF mode [JXJA08].

The parameters that will be used in Section 5.3 are listed as follows:

y_{br}: The received signal at *Relay r* from *BS b* (*BS b to Relay r* link).

y_{bu}: The received signal at *User u* from *BS b* (*BS b to User u* link).

y_{ru}: The received signal at *User u* from *Relay r* (*Relay r to User u* link).

a_{br} : Channel gain from *BS b* to *Relay r*.

a_{bu} : Channel gain from *BS b* to *User u*.

a_{ru} : Channel gain from *Relay r* to *User u*. $|a_{ru}|^2$ is the channel power gain from *Relay r* to *User u*.

λ_r: Amplified factor of *Relay r*.

P_b: Transmit power of *BS b*.

P_r: Transmit power of *Relay r*.

$I_r[n]$: Inter-cell interference at *Relay r* at TS *n*.

$I_u[n]$: Inter-cell interference at *User u* at TS *n*.

$C_u^{AF\,r}$: Achievable rate of *User u* with *Relay r* assistance.

$C_u^{NO\ AF}$: Achievable rate of *User u* without relay assistance.

C_r^{BS}: *Relay r*'s achievable rate, with the same number of subcarriers being allocated to *Relay r*. C_r^{BS} reflects *Relay r*'s total rate loss.

x_b : Useful signal from *BS b*.

σ^2 : Common variance of the Gaussian white noise.

B: Bandwidth in the user relaying model.

5.3.3 Analysis of User Relaying Model

Based on the user relaying model, this section analyses the achievable rate of shifted *User u*. Besides, the impact of energy consumption of *Relay r* is also discussed.

At TS *n,* the received signals at *User u* and *Relay r* are given by (5.14) and (5.15), respectively.

$$y_{bu}[n] = a_{bu}x_b[n] + Z_u[n] + I_u[n] \tag{5.14}$$

$$y_{br}[n] = a_{br}x_b[n] + Z_r[n] + I_r[n] \tag{5.15}$$

where $Z_u[n]$ and $Z_r[n]$ are the noise at *User u* and *Relay r*, respectively. $I_u[n]$ and $I_r[n]$ are the inter-cell interference at *User u* and *Relay r*, respectively. a_{bu} is the channel gain from *BS b* to *User u*. a_{br} is the channel gain from *BS b* to *Relay r*.

5.3.3.1 Achievable Rate of *User u* in User Relaying Model

In the AF mode [JXJA08], *Relay r* amplifies all received signals and forwards to the shifted *User u* at TS *n+1*. From (5.15), the amplified factor of *Relay r* can be expressed as λ_r, using (5.17).

$$\lambda_r^2(|a_{br}|^2 P_b + \sigma^2 + |I_r[n]|^2) = P_r \tag{5.16}$$

$$\Rightarrow \quad \lambda_r^2 = \frac{P_r}{|a_{br}|^2 P_b + \sigma^2 + |I_r[n]|^2} \tag{5.17}$$

where P_b and P_r are the transmit power of *BS b* and *Relay r*, respectively. σ^2 is the common variance of the Gaussian white noise. $|\ |$ denotes the amplitude of the symbol. For example, $|I_r[n]|^2$ is the interference power at *Relay r*.

At TS *n+1*, the received signals at *User u* from *Relay r* and from *BS b* are discussed in *i)* and *ii)*.

i) At TS *n+1*, the received signal at *User u* from *Relay r* (*Link* L_{ru} in Figure 5.17) can be calculated as

$$y_{ru}[n+1] = \lambda_r y_{br}[n] a_{ru} + Z_u[n+1] + I_u[n+1] \tag{5.18a}$$

$$= \lambda_r (a_{br} x_b[n] + Z_r[n] + I_r[n]) a_{ru} + Z_u[n+1] + I_u[n+1] \tag{5.18b}$$

$$= (\lambda_r a_{ru} a_{br} x_b[n]) + (\lambda_r a_{ru} Z_r[n] + Z_u[n+1]) + \lambda_r a_{ru} I_r[n] + I_u[n+1] \tag{5.18c}$$

where $y_{br}[n]$ refers to (5.15), a_{ru} is the channel gain from *Relay r* to *User u*.

ii) At TS *n+1*, the received signal at *User u* from *BS b* (*Link* L_{bu} in Figure 5.17) is denoted as $y_{bu}[n+1]$, using (5.19). From (5.19), SINR of *User u* at TS *n+1* from L_{bu} is denoted as $SINR_{u,n+1}^{(L_{bu})}$, using (5.20).

$$y_{bu}[n+1] = a_{bu} x_b[n+1] + Z_u[n+1] + I_u[n+1] \tag{5.19}$$

$$SINR_{u,n+1}^{(L_{bu})} = \frac{P_b \mid a_{bu} \mid^2}{|I_u[n+1]|^2 + \sigma^2} \tag{5.20}$$

iii) At TS *n*, the received signal at *User u* from *BS b* (*Link* L_{bu} in Figure 5.17) is shown in (5.14). Therefore, SINR of *User u* at TS slot *n* from L_{bu} can be expressed as

$$SINR_{u,n}^{(L_{bu})} = \frac{P_b \mid a_{bu} \mid^2}{|I_u[n]|^2 + \sigma^2} \tag{5.21}$$

As introduced in the user relaying model, $x_b[n]$ and $x_b[n+1]$ are the identical signal and transmitted in three separate links. Specifically, *User u* receives $x_b[n]$ from *BS b* directly at TS *n*, as depicted in (5.14); *User u* also receives $x_b[n]$ forwarded by *Relay r* at TS *n+1*, as depicted in (5.18); in addition, *User u* receives $x_b[n+1]$ from *BS b* directly at TS *n+1*, as depicted in (5.19). *User u* combines the signal from three separate links to enhance the signal quality. The estimated SINR of *User u* is

$$SINR_u^{AF\,r} \approx \frac{|a_{bu}|^2 P_b}{\sigma^2 + |I_u[n]|^2} + \frac{|a_{bu}|^2 P_b}{\sigma^2 + |I_u[n+1]|^2} + \frac{\lambda_r^2 |a_{ru}|^2 |a_{br}|^2 P_b}{\lambda_r^2 |a_{ru}|^2 (\sigma^2 + |I_r[n]|^2) + \sigma^2 + |I_u[n+1]|^2} \tag{5.22}$$

Equation (5.22) is the estimated SINR, which is used to select suitable relay. Then, the estimated achievable rate of *User u* with *Relay r* assistance is

$$C_u^{AF\,r} \approx \frac{B}{2} \log_2(1 + SINR_u^{AF\,r}) \tag{5.23}$$

where *B* is the bandwidth, $\frac{1}{2}$ denotes that *User u* receives the identical signal at two consecutive time slots [WTJLHL10].

Relay selection impacts the value of $|a_{ru}|^2$, $|a_{br}|^2$ and $|I_r[n]|^2$. From Equation (5.23), selecting a suitable relay user can improve the achievable rate of the shifted user.

5.3.3.2 Achievable Rate of *User u* without Relay

If there is no relay link, *User u* only receives signal from *BS b* (*Link* L_{bu} in Figure 5.17) at TS *n*

and TS *n+1*. From (5.20) (5.21), the achievable rate of *User u* without relay is

$$C_u^{NO\ AF} = \frac{B}{2}\log_2(1+\frac{P_b\,|\,a_{bu}\,|^2}{|I_u[n]|^2+\sigma^2}+\frac{P_b\,|\,a_{bu}\,|^2}{|I_u[n+1]|^2+\sigma^2}) \tag{5.24}$$

5.3.3.3 Total Rate Loss of *Relay r*

In the user relaying model, *Relay r* amplifies signal power and forwards signal to *User u* at TS *n+1*. This consumes the energy of *Relay r* and shortens *Relay r* battery working time, which will result in the total rate loss of *Relay r* during battery working time. We define C_r^{BS} as *Relay r*'s achievable rate, with the same number of subcarriers (the same bandwidth) being allocated to *Relay r*. Hence, C_r^{BS} reflects *Relay r*'s total rate loss, and C_r^{BS} indicates the impact on the energy consumption of *Relay r*.

If *Relay r* is an active user, the received signal at *Relay r* at TS *n+1* is given by (5.25). Correspondingly, the achievable SINR of *Relay r* at TS *n+1* is defined as $SINR_{r,n+1}$, using (5.26).

$$y_{br}[n+1] = a_{br}x_b[n+1] + Z_r[n+1] + I_r[n+1] \tag{5.25}$$

$$SINR_{r,n+1} = \frac{P_b\,|\,a_{br}\,|^2}{|I_r[n+1]|^2+\sigma^2} \tag{5.26}$$

where P_b is the transmit power of *BS b*, $|I_r[n+1]|^2$ is the interference power at *Relay r* at TS *n+1*, $|a_{br}|^2$ is the channel power gain from *BS b* to *Relay r*. From (5.26), C_r^{BS} can be calculated as

$$C_r^{BS} = \frac{B}{2}\log_2(1+\frac{P_b\,|\,a_{br}\,|^2}{|I_r[n+1]|^2+\sigma^2}) \tag{5.27}$$

Relay selection impacts the value of $|a_{br}|^2$ and $|I_r[n+1]|^2$. Equation (5.27) indicates that selecting an appropriate relay can reduce the total rate loss of the relay user.

5.3.4 Proposed URTS Scheme

From the analysis above, the user relaying model provides a relay link to improve the achievable rate of the shifted user. However, this model also consumes the battery power of the relay user and shortens the relay user's working time, which will reduce the relay user's total rate. Both the factor of the shifted user's achievable rate and the factor of the relay user's total rate loss should be considered jointly in relay selection.

Therefore, based on the user relaying model, URTS scheme is designed. The key of URTS scheme lies in designing a utility function to select an appropriate relay for the trade-off between shifted user's performance and the impact of relay user's energy consumption.

5.3.4.1 Weight of Traffic Shifting

In order to select a suitable *Relay r* to increase the achievable rate of the shifted *User u*, the weight of traffic shifting (WTS) is designed as $\Psi_{r,WTS}$. As shown in (5.28), $\Psi_{r,WTS}$ equals the ratio of *User u's* achievable rate with *Relay r* assistance $C_u^{AF\,r}$ to *User u's* achievable rate without relay $C_u^{NO\ AF}$. Therefore, $\Psi_{r,WTS}$ indicates the achievable rate gain of *User u*.

$$\psi_{r,WTS} = \frac{C_u^{AF\,r}}{C_u^{NO\ AF}} \qquad r \in \{1,2...R\} \tag{5.28}$$

5.3.4.2 Weight of Energy Consumption

The energy consumption of *Relay r* shortens battery working time and reduces the total rate of *Relay r*. Under the similar energy consumption of the non-active *Relay r*, the weight of energy consumption (WEC) is designed as $\Psi_{r,WEC}$. WEC compares two impact of energy consumption, including the total rate loss of *Relay r* (C_r^{BS}), the rate improvement of *User u* with *Relay r* assistance ($C_u^{AF\,r} - C_u^{NO\ AF}$). $\Psi_{r,WEC}$ is calculated as

$$\psi_{r,WEC} = \frac{C_r^{BS}}{C_u^{AF\,r} - C_u^{NO\ AF}} \qquad r \in \{1,2...R\} \tag{5.29}$$

$\Psi_{r,WEC}$ indicates the impact of energy consumption of *Relay r*. In (5.29), the higher total rate loss of *Relay r* leads to the higher $\Psi_{r,WEC}$.

5.3.4.3 Utility Function based Relay Selection

In order to select a suitable relay to reach the trade-off between the shifted user's performance and relay user's energy consumption, a utility function u_r is designed as

$$u_r = \frac{\psi_{r,WTS}}{\psi_{r,WEC}} = \frac{C_u^{AF\,r} \times (C_u^{AF\,r} - C_u^{NO\,AF})}{C_u^{NO\,AF} \times C_r^{BS}} \qquad r \in \{1,2...R\} \tag{5.30}$$

According to (5.30), the *higher User u's achievable rate with Relay r assistance* can lead to higher u_r. Meanwhile, the *lower total rate loss of Relay r* can lead to higher u_r. Hence, URTS scheme tries to select *Relay k* to maximize u_r, as

$$Relay\ k = \arg\max_r\ u_r \tag{5.31a}$$

$$\Longrightarrow \quad Relay\ k = \arg\max_r \frac{C_u^{AF\,r} \times (C_u^{AF\,r} - C_u^{NO\,AF})}{C_u^{NO\,AF} \times C_r^{BS}} \tag{5.31b}$$

From (5.31), the utility function relates to $C_u^{AF\,r}$, C_r^{BS} and $C_u^{NO\,AF}$. *User u* has a correspondingly fixed $C_u^{NO\,AF}$, given by (5.24).

Both $C_u^{AF\,r}$ and C_r^{BS} vary with different *Relay r*. From (5.23), $C_u^{AF\,r}$ is based on three variable parameters: $|a_{ru}|^2$, $|a_{br}|^2$, and $|I_r[n]|^2$. From (5.27), C_r^{BS} is based on $|a_{br}|^2$ and $|I_r[n+1]|^2$.

5.3.4.4 URTS Scheme Process

From Section 5.3.4.3, *User u* can calculate the utility function to choose suitable relay, only under knowing the value of $|a_{br}|^2$, $|a_{ru}|^2$, $|I_r[n]|^2$ and $|I_r[n+1]|^2$. In order to reduce the complexity and the signalling load, the URTS scheme calculates them according to existing/measurable parameters in other RRM functionalities, e.g., cell selection, admission control. Specifically, they can be estimated as:

- $|a_{br}|^2$ (channel power gain from *BS b* to *Relay r)*: Since *Relay r* knows the received RSRP from *BS b*, as well as *BS b*'s transmit power (which can be informed from *BS b* via control channel), *Relay r* estimates $|a_{br}|^2$ as

$$|a_{br}|^2 \approx \frac{\textit{Relay r's received RSRP from BS b}}{\textit{BS b's transmit power } P_b} \tag{5.32}$$

- $|a_{ru}|^2$ (channel power gain from *Relay r* to *User u)*: After *Relay r* responding to *User u*, *User u* knows the received power of *Relay r's* response signal. Besides, *Relay r* reports P_r to *User u* in URTS scheme, as shown in Figure 5.18). *User u* calculates $|a_{ru}|^2$ as

$$|a_{ru}|^2 \approx \frac{\textit{User u's received power of Relay r's response signal}}{\textit{Relay r's transmit power } P_r} \tag{5.33}$$

- $|I_r[n]|^2$, $|I_r[n+1]|^2$ (interference power at *Relay r,* at TS *n* and TS *n+1*): In the full frequency reuse OFDMA cellular networks, precise interference estimation is difficult. It is because *Relay r's* interference, which is imposed by other cells using the co-channel subcarriers, is varying due to the dynamic subcarriers allocation of neighbouring cells. To reduce the estimation complexity, *Relay r* considers the RSRP from all neighbouring BSs as the interference, and then calculates the theoretically heaviest interference $|I_r|^2$.

Similarly, *User u* estimates $|a_{bu}|^2$ and $|I_u|^2$, which are not varying with different *Relay r*. The flowchart of the URTS scheme is shown in Figure 5.18, which involves the process of shifted *User u* and *Relay r.*

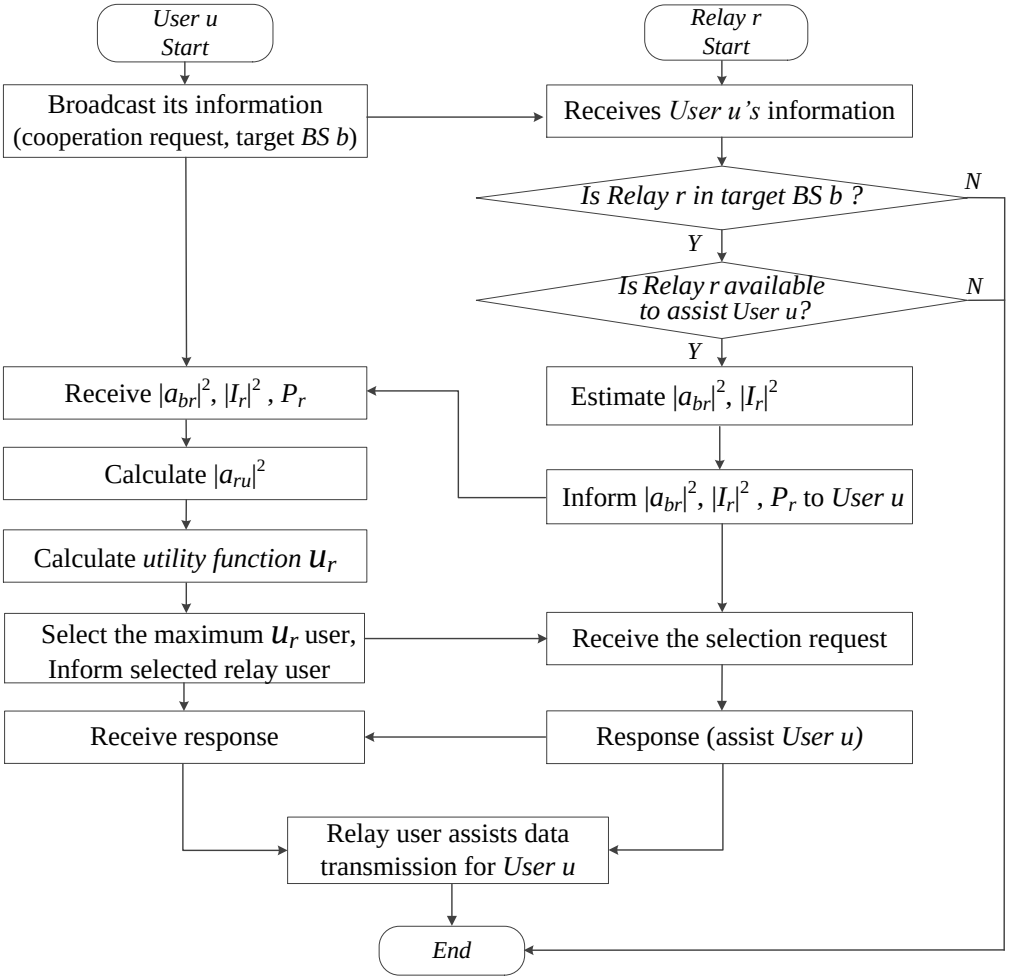


Figure 5.18 Flowchart of URTS scheme

As shown in Figure 5.18, if a user in the hot-spot BS needs to be shifted to the target *BS b*, the shifted *User u* broadcasts the message of cooperation request and target *BS ID*, namely *BS b*.

After receiving the broadcast message, the non-active user judges whether it is in the coverage of *BS b* and whether it is available to assist *User u*. This is because that a non-active user in the coverage of *BS b* can only assist a shifted user at a time, in order to reduce the processing complexity. If it is, the non-active *Relay r* calculates $|a_{br}|^2$ from Equation (5.32). Besides, *Relay r* estimates $|I_r|^2$ as the sum of RSRP from all neighbouring BSs. Then *Relay r* responds and sends messages of $|a_{br}|^2$, $|I_r|^2$ and P_r to *User u*.

After receiving the responses, *User u* calculates $|a_{ru}|^2$ from Equation (5.33). In addition, *User u* estimates its corresponding $|a_{bu}|^2$ and $|I_u|^2$. Then *User u* estimates $C_u^{AF\,r}$, C_r^{BS} and $C_u^{NO\,AF}$. Based on the estimated $C_u^{AF\,r}$, C_r^{BS} and $C_u^{NO\,AF}$, *User u* calculates the utility function u_r / s of all responding non-active users. Then *User u* selects a non-active user, which has the largest

u_r as the relay user.

After the relay user selection, the selected non-active user forwards signal to *User u*.

Note that multiple shifted users may request one non-active user at the same time. Under this scenario, the non-active user chooses one shift user, from which the non-active user receives the strongest broadcast power. This is because that the high received power from the shifted user indicates good link quality between the two users.

5.3.5 Performance Analysis

5.3.5.1 Simulation Schemes Introduction

The proposed scheme is evaluated by the system-level simulation platform designed in Chapter 3. The key simulation parameters are introduced in Section 3.7.2. This simulator generates both active users and non-active users in the simulation area. The two types of users have the similar distribution: 70% active users (green circle) and 70% non-active users (blue dot) are located in three hot-spot areas, as shown in Figure 5.19. Besides, four schemes are simulated.

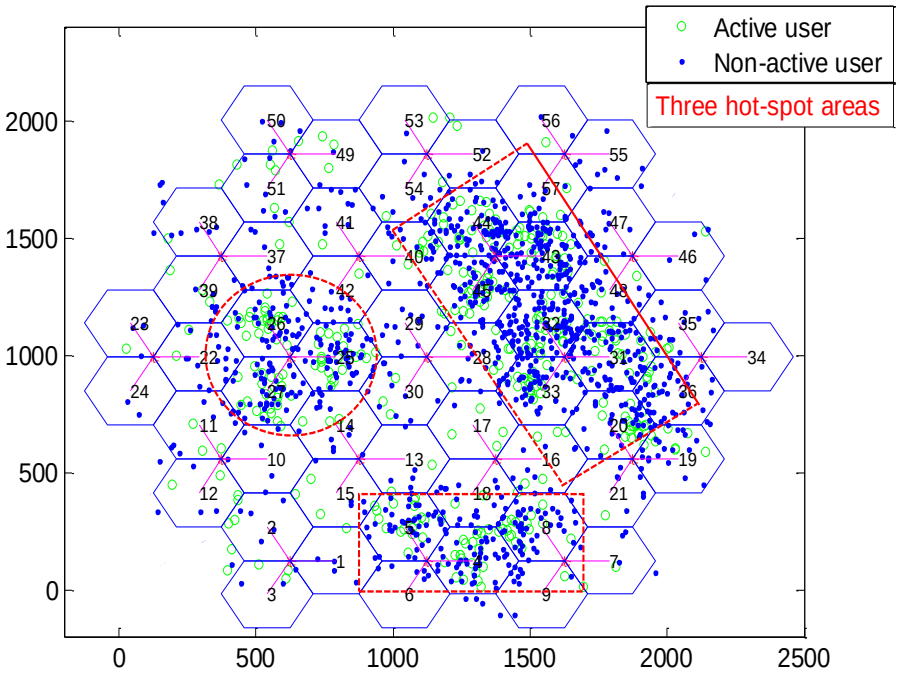


Figure 5.19 Simulation scenario for URTS scheme

1) URTS (called CCLB with utility function user relaying scheme in Figure 5.21 - Figure 5.24)

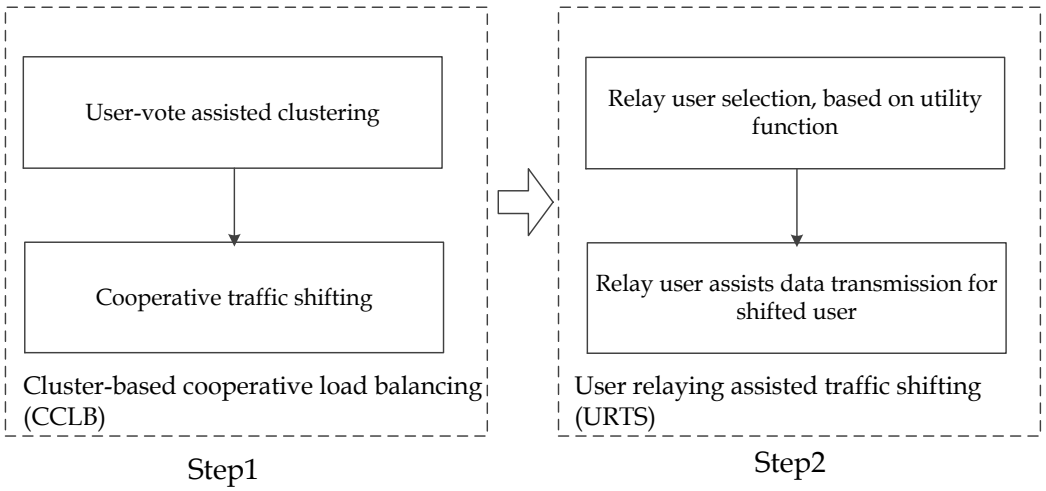


Figure 5.20 Overall simulation flowchart of URTS scheme

The proposed (utility function based) URTS scheme is simulated. Figure 5.20 shows the overall simulation flowchart. A hot-spot cell employs the CCLB scheme to shift users to partner cells. Then the shifted user employs the proposed URTS scheme for transmission. Specifically, the shifted user calculates the utility function of non-active users, in order to choose a non-active user with the largest value of utility function as relay user. Finally, the relay user forwards signal for the shifted user.

2) CCLB scheme

This simulator also simulates the standalone CCLB scheme (without user relaying), which is previously proposed in Chapter 4. In CCLB scheme, cluster head adjusts HO_{off} towards partner $Cell_b$. Then *User u* in the cluster head will be shifted to $Cell_b$ without relay assistance.

3) Typical MLB scheme

The typical MLB in [KAPTK10] is simulated as the benchmarking scheme for comparison. This scheme is introduced in Section 2.4.3.2.

4) WTS user relaying scheme (called CCLB with WTS user relaying scheme in Figure 5.21-Figure 5.24)

In order to evaluate the performance improvement by adopting the proposed utility

function, the reference CCLB with WTS user relaying scheme is simulated (Note that this is also our proposed scheme). The simulation flow is similar to Figure 5.20. The difference is that in CCLB with WTS user relaying scheme, a shifted user only considers the *WTS* (weight of traffic shifting) during the relay selection. As discussed in Section 5.3.4.1, CCLB with WTS user relaying scheme aims at selecting the relay which can best improve the achievable rate of the shifted user, while *WTS* does not consider the total rate loss of the relay user.

5.3.5.2 Simulation Results

A shifted user may suffer load balancing handover failure when the partner cell cannot provide sufficient subcarriers to meet the shifted user's service requirement, or the partner cell cannot provide the shifted user with the required SINR to sustain connection [JBTMK10] [SOCRATES10]. In Figure 5.21, the *CCLB* scheme has lower load balancing handover rate than the typical MLB scheme [KAPTK10], because *CCLB* scheme can address the heavily loaded public partner. Compared with the *CCLB* scheme, the proposed *CCLB with utility function user relaying* scheme can further reduce the handover failure rate. This is because that the relay link can enhance the link quality of the shifted user, and then the handover failure rate is reduced. For example, under 600 users scenario, the handover failure in the proposed *CCLB with utility function user relaying* scheme *is* 2.3%, compared with 9% of *CCLB* scheme and 18% of typical MLB scheme.

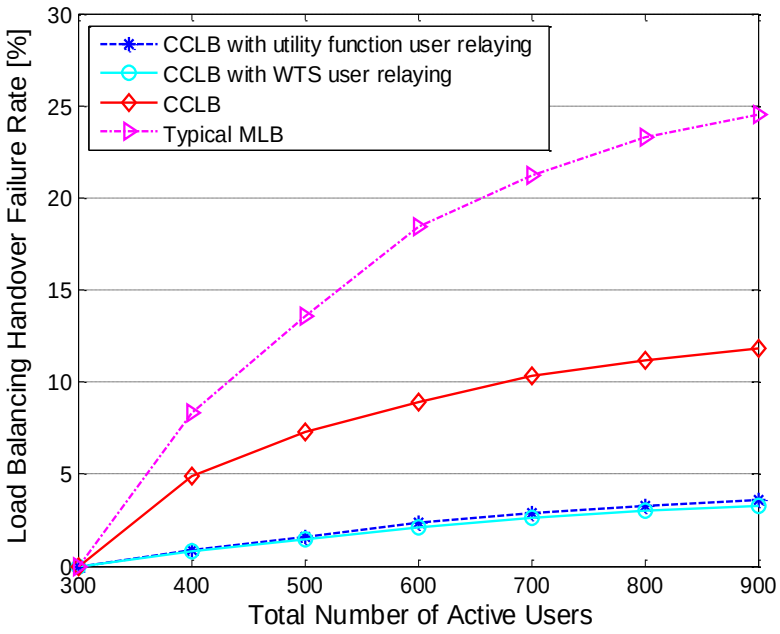


Figure 5.21 Load balancing handover failure rate comparison

Figure 5.22 further evaluates the proposed *CCLB with utility function user relaying* scheme. In order to examine its performance for assisting shifted users of different link qualities, four categories of shifted users are considered according to their SINR. The four categories encompass: SINR lower than 1; SINR between 1 and 2; SINR between 2 and 6; SINR between 6 and 12.

Among four categories, the poor link quality shifted users (SINR<1, 1<SINR<2), experience large SINR improvement via *CCLB with utility function user relaying* scheme. The proposed scheme can increase nearly 75% SINR for shifted users in SINR<1 and 1<SINR<2 categories. The proposed scheme also effectively improves SINR for the medium link quality shifted users, e.g., 32% SINR improvement in for shifted users in 2<SINR<6 category. The proposed scheme can also increase the SINR of good link quality shifted users, e.g., 6<SINR<12 category. But their SINR enhancements are not as outstanding as poor/medium link quality users. For example, shifted users in 6<SINR<12 category experience 20% SINR increase.

Due to shifted users' improved SINR and the reduced load balancing handover failure rate, Figure 5.23 shows that the *CCLB with utility function user relaying* scheme can improve the overall rate of all shifted users, compared with *CCLB* scheme.

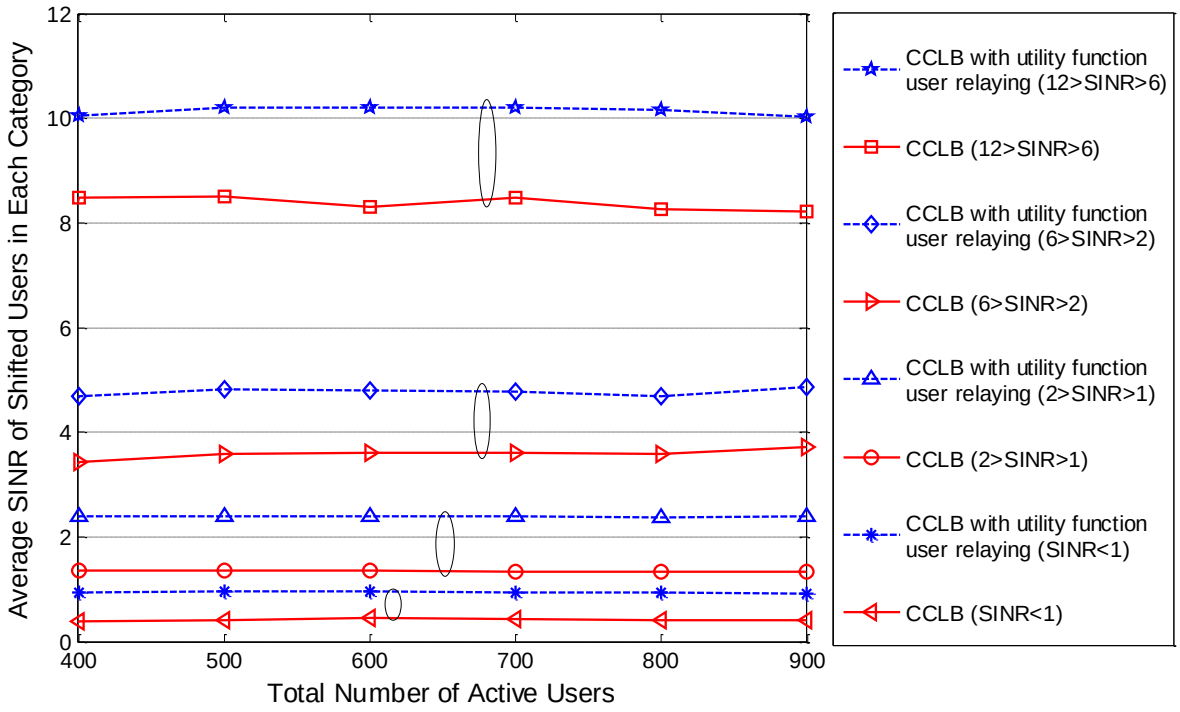


Figure 5.22 SINR comparison of shifted users in different SINR categories

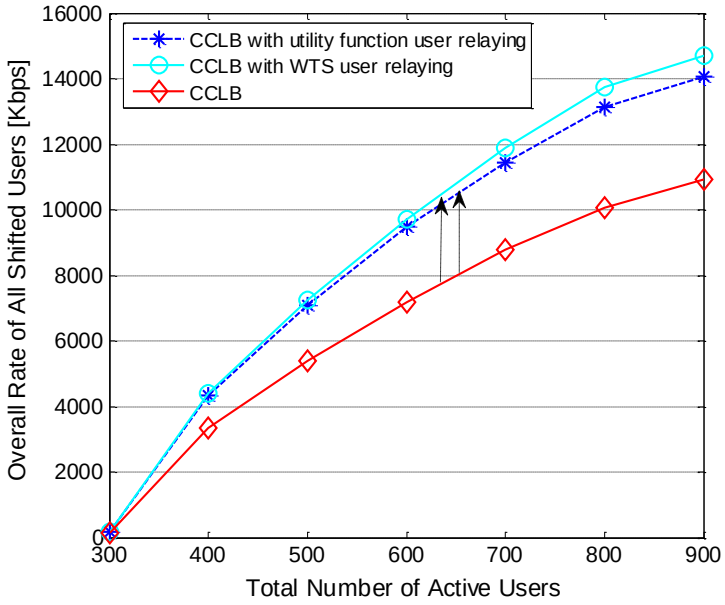


Figure 5.23 Comparison of overall rate of all shifted users

Furthermore, the trade-off between shifted users performance and relay users performance is demonstrated in Figure 5.21, Figure 5.23 and Figure 5.24.

As depicted in Figure 5.21, the *CCLB with utility function user relaying* scheme can reach a similar performance of load balancing handover failure rate, compared with the *CCLB with WTS user relaying* scheme.

Figure 5.23 shows the overall rate of all shifted users. Both the *CCLB with utility function user relaying* scheme and *CCLB with WTS user relaying* scheme can effectively improve the total rate of shifted users. For example, compared with *CCLB* scheme, the *CCLB with utility function user relaying* scheme can increase the rate by 31%, and the *CCLB with WTS user relaying* scheme can increase the rate by 35%, under 700 users scenario. The reason of the slight difference is that the *WTS* scheme only considers the rate improvement of the shifted user, while the utility function also considers the rate loss of the relay user.

Figure 5.24 depicts the overall rate loss of all relay users. The overall rate loss in the *CCLB with utility function user relaying* scheme is nearly 23%~30% less than that in the *CCLB with WTS user relaying* scheme.

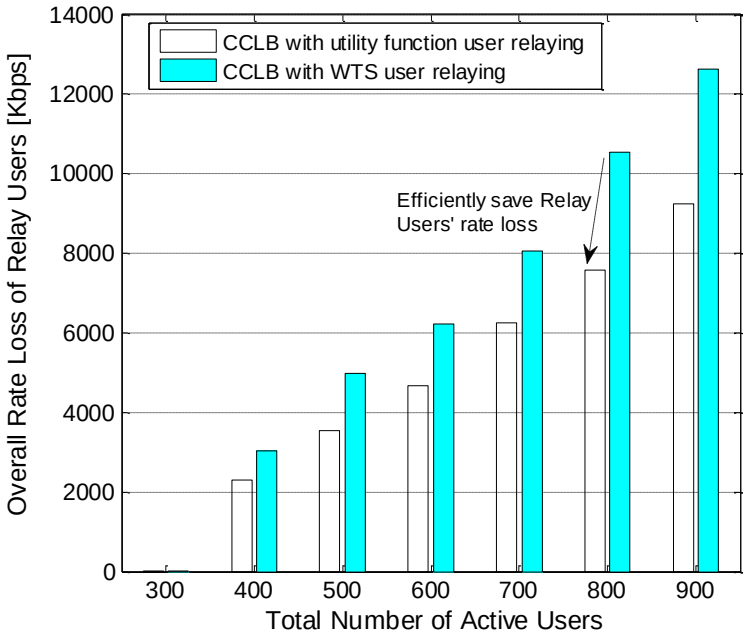


Figure 5.24 Comparison of overall rate loss of all relay users

From the analysis above, both *CCLB with utility function user relaying* scheme and *CCLB with WTS user relaying* scheme bring similar performance for shifted users. Meanwhile, *CCLB with utility function user relaying* scheme can effectively reduce the rate loss of relay users. Therefore, the proposed utility function can reach a good trade-off between shifted users' performance and relay users' performance.

5.4 Summary

In Section 5.2, CCLB scheme is modified to be feasible in fixed relay cellular networks. It includes three stages: user-vote assisted clustering to deal with the virtual partner problem in fixed relay networks; cell-level cooperative traffic shifting algorithm to deal with the aggravating load problem of public partner; after above two stages, a relay-level user shifting algorithm is designed particularly for fixed relay cellular networks. The relay-level user shifting algorithm can address the RS aggravating load problem. Simulation results show that the proposed scheme can select a small number of partners to effectively shift traffic. The scheme can keep the public partner's load lower than the heavily loaded threshold. In addition, the relay-level user shifting algorithm can mitigate the heavily loaded RS and effectively reduce the load balancing handover failure rate.

In Section 5.3, the shifted user's link quality degradation problem is discussed. This problem

is confronted in cellular networks without fixed relay deployment. In Section 5.3, a user relaying model is employed to enhance the link quality of shifted users. Furthermore, based on the user relaying model, a user relaying assisted traffic shifting scheme (URTS) is designed. The URTS scheme can effectively enhance the link quality of shifted users under low cost of relay users' energy consumption. Simulation results show that the URTS scheme can increase the SINR of shifted user by 20%~75% and reduce the load balancing handover failure rate. Moreover, the utility function based relay user selection mechanism can reach a good trade-off between shifted user's performance improvement and relay user's rate loss.

Chapter 6 Conclusions and Future Work

This thesis has proposed a self-organising cluster-based cooperative load balancing (CCLB) scheme for single-hop cellular networks. Then the CCLB scheme has been modified to apply in multi-hop cellular networks with relay deployed. The aim of the CCLB scheme is to balance load among cells and deal with problems confronted in conventional MLB schemes.

6.1 Specific Conclusions

As introduced in Chapter 4, the proposed CCLB scheme consists of two stages: partner selection and traffic shifting. Correspondingly, the CCLB scheme includes following specific outcomes:

- In the partner selection stage, a user-vote assisted clustering algorithm has been proposed in Section 4.4. This algorithm considers users' channel condition and neighbouring cells' load. Compared with the conventional pure load based partner selection, the proposed clustering algorithm can select more suitable partners and mitigate the virtual partner problem. Using the proposed algorithm, Section 4.6.1 demonstrates that two best partners can reach a good load balancing performance and reduce the number of HO_{off} adjustments by nearly 60%.

- In the cell-level traffic shifting stage, a cooperative traffic shifting algorithm has been proposed in Section 4.5. This algorithm includes inter-cluster cooperation and intra-cluster cooperation. During the inter-cluster cooperation, relative load response model (RLRM) coordinates multiple cluster heads' traffic shifting requests to one public partner. RLRM can mitigate the aggravating load problem of public partner. The intra-cluster cooperation mechanism can reduce cluster head's load and minimise the average call

blocking probabilities of cluster head's partners.

As introduced in Section 5.2, the CCLB scheme has been modified to apply in fixed relay cellular networks. The specific outcomes include:

- Since most shifted users are served by RS of neighbouring cells, the user-vote assisted clustering algorithm considers user channel condition from RS and neighbouring cells' load. Section 5.2.7 shows that this algorithm can select a small number of partners to effectively balance load and address virtual partner problem.

- A relay-level user shifting algorithm has been designed for fixed relay cellular networks. This algorithm considers the users' channel condition and analyses RS's spectrum usage. Section 5.2.7 shows that this algorithm can mitigate RS aggravating load problem and reduce the load balancing handover failure rate by nearly 20%.

As introduced in Section 5.3, after the CCLB scheme implementation, a user relaying assisted traffic shifting (URTS) scheme has been proposed.

- The URTS scheme employs the non-active users as the mobile relay to transmit data for shifted users. This scheme can address the link quality degradation of shifted users.

- Compared with conventional MLB, Section 5.3.5 shows that the URTS scheme can improve the shifted user's SINR and reduce the load balancing handover failure rate. In addition, the URTS scheme can reach a good trade-off between shifted user's performance improvement and relay user's rate loss.

6.2 Future Work

In this thesis, the focus has been on partner selection stage and traffic shifting stage in mobility load balancing. The CCLB scheme has been proposed to deal with virtual partner problem and aggravating load problem. Besides, single-hop and multi-hop cellular networks are considered. In the future work, some issues and technologies will be researched:

- Uplink and downlink joint optimisation: It is generally known that load balancing is always triggered in downlink cellular networks, due to the asymmetric service between uplink and downlink. After load balancing, the cellular networks might suffer the

inconsistent cell coverage between downlink and uplink. Therefore, the future work involves the uplink and downlink joint optimisation to improve the QoS in uplink and balance traffic load in downlink.

- Research of the interaction between handover and mobility load balancing: Mobility load balancing aims at shifting edge users to lightly loaded neighbouring cells via HO_{off} adjustment. The adjusted HO_{off} may impact the handover performance, e.g., handover failure. Hence, the future work involves the interaction between handover and mobility load balancing. Based on this research, we can further modify the CCLB scheme to improve the handover performance.

- Induced admission control algorithm: Besides load balancing functionality, other RRM functionality, such as admission control, can also control hot-spot cell's load. Specifically, we will design an induced admission control algorithm. In this novel algorithm, a hot-spot cell does not reject a new call user's admission request, instead, this hot-spot cell sends induction information to help the new call user access to a suitable lightly loaded neighbouring cell.

- Research of load balancing in multi-tier and multi-RAT networks: There are other techniques that can balance load, particularly the traffic shifting in multi-tier and multi-RAT networks, e.g., femtocell and micro-cell. One line of approach for further work can be on applying the basic idea of CCLB scheme to cooperatively balance load in multi-tier and multi-RAT networks. For example, the method of user-vote assisted clustering can be considered in tier/RAT selection.

References

[3GPP08a] 3GPP TS 36.211 V8.2.0, "E-UTRA physical channels and modulation (Release 8)," March 2008.

[3GPP08b] 3GPP TR 36.902 V1.0.0, "Self-configuring and self-optimising network use cases and solutions (Release 8)," December 2008.

[3GPP08c] 3GPP TR 25.996 V8.0.0, "Spatial channel model for multiple input multiple output (MIMO) simulations (Release 8)," December 2008.

[3GPP09] 3GPP TR 25.816 V8.0.0, "UMTS 900 MHz work item technical report (Release 8)," September 2009.

[3GPP10a] 3GPP TS 36.423 V10.0.0, "X2 application protocol (X2AP) (Release 10)," December 2010.

[3GPP10b] 3GPP TS 36.314 V10.0.0, "Layer 2 - measurements (Release 10)," December 2010.

[3GPP10c] 3GPP TS 36.300 V9.5.0, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN) overall description (Release 9)," September 2010.

[3GPP10d] 3GPP TR 36.814 V9.0.0, "Further advancements for E-UTRA physical layer aspects (Release 9)," March 2010.

[3GPP10e] 3GPP TR 36.806 V9.0.0, "Relay architectures for E-UTRA (LTE-Advanced) (Release 9)," March 2010.

[3GPP11a] 3GPP TS 36.420 V10.0.1, "X2 general aspects and principles (Release 10)," March 2011.

[3GPP11b] 3GPP TR 36.902 V9.3.1, "Self-configuring and self-optimizing network (SON) use cases and solutions (Release 9)," March 2011.

[3GPP11c] 3GPP TR 25.942 V10.0.0, "Radio frequency (RF) system scenarios (Release 10)," April 2011.

[3GPP11d] 3GPP TS 36.304 V9.9.0, "User equipment (UE) procedures in idle mode (Release 9) ," December 2011

[3GPP12] 3GPP TS 36.331 V11.1.0, "Radio resource control (RRC) protocol specification (Release 11)," September 2012.

[ANHV10] Mehdi Alasti, Behnam Neekzad, Jie Hui, and Rath Vannithamby, "Quality of service in WiMAX and LTE networks," *IEEE Communications Magazine*, vol.48, no.5, pp.104-111, May 2010.

[ACMRP07] Mohmmad Anas, Francesco D. Calabrese, Preben E. Mogensen, Claudio Rosa, and Klaus I. Pedersen, "Performance evaluation of received signal strength based hard handover for UTRAN LTE," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.1046-1050, April 2007.

[Bertsekas99] Dimitri P. Bertsekas, "Nonlinear programming," Athena Scientific, 1999.

[CA09] Francisco Rodrigo Porto Cavalcanti and Soren Andersson, "Optimising wireless communication systems," Springer, 2009.

[Chen03] Yue Chen, "Soft handover issues in radio resource management for 3G WCDMA networks," PhD Thesis, Queen Mary, University of London, UK, September 2003.

[CJC09] Sunghyun Cho, Edward W. Jang, and John M. Cioffi, "Handover in multihop cellular networks," *IEEE Communications Magazine*, vol.47, no.7, pp.64-73, July 2009.

[CYOCY08] Basak Can, Halim Yanikomeroglu, Furuzan Atay Onat, Elisabeth De Carvalho, and Hiroyuki Yomo, "Efficient cooperative diversity schemes and radio resource allocation for IEEE 802.16j," *IEEE Wireless Communications and Networking Conference (WCNC)*, pp.36-41, April 3 2008.

[DS10] Ke-Lin Du and M. N. S. Swamy, "Wireless communication systems: from RF subsystems to 4G enabling technologies," Cambridge University Press, 2010.

[DSJ97] Sajal K. Das, Sanjoy K. Sen, and Rajeev Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment," *Wireless Networks*, vol.3, pp.333 -347, 1997.

[Damosso98] Eraldo Damosso, "Digital mobile radio towards future generation systems," COST 231 Final Report, 1998.

[E³08] FP7 End-to-End Efficiency (E³), Eckard Bogenfeld and Ingo Gaspard, "Self-x in radio access networks," E³ White Paper, V1.0, December 2008, https://ict-e3.eu/project/white_papers/whitepapers.html/

[EP73] Joel S. Engel and Martin M. Peritsky, "Statistically-optimum dynamic sever assignment in systems with interfering severs," *IEEE Transactions on Communications*, vol.21, no.11, pp.1287-1293, November 1973.

[FHHX09] Jiang Fan, Tian Hui, Xu Haibo, and Wang Xijun, "A novel relay based load balancing scheme and performance analysis using Markov models," *IEEE Vehicular Technology Conference-Spring (VTC-Spring),* April 2009.

[FSCK10] Minghai Feng, Xiaoming She, Lan Chen, and Yoshihisa Kishiyama, "Enhanced dynamic cell selection with muting scheme for DL CoMP in LTE-A," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.1-5, May 2010.

[FW11] Jiang Fan and Benchao Wang, "A load balancing relay selection algorithm for relay based cellular networks," *International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, September 2011.

[GKTH97] Aura Ganz, C.M. Krishna, Dingyi Tang, and Zygmunt J. Haas, "On optimal design of multitier wireless cellular systems," *IEEE Communications Magazine,* vol.35, no.2, pp.88-93, February 1997.

[Goldsmith05] Andrea Goldsmith, "Wireless communications," Cambridge University Press: New York, 2005.

[GZLLZ07] Lei Guan, Jianhua Zhang, Jianing Li, Guangyi Liu, and Ping Zhang, "Spectral efficient frequency allocation scheme in multihop cellular network," *IEEE Vehicular Technology Conference-Fall (VTC-Fall)*, pp. 1446-1450, September 2007.

[Hanson99] Morgan A. Hanson, "Invexity and the Kuhn-Tucker theorem," *Journal of Mathematical Analysis and Applications*, vol.236, pp.594–604, 1999.

[HCLL12] Ren-Hung Hwang, Ben-Jye Chang, Yan-Min Lin, and Ying-Hsin Liang, "Adaptive load-balancing association handoff approach for increasing utilization and improving GoS in mobile WiMAX networks," *Wiley Wireless Communications and Mobile Computing* 2012, Wiley Online Library.

[HGV10] Angela Hernández, Isreal Guío, and Antonio Valdovinos, "Radio resource allocation for interference management in mobile broadband OFDMA based networks," *Wiley Wireless Communications and Mobile Computing*, vol.10, no.11,

pp.1409–1430, November 2010.

[HT04] Harri Holma and Antti Toskala, "WCDMA for UMTS: radio access for third generation mobile communications," John Wiley & Sons, 2004.

[HT09] Harri Holma and Antti Toskala, "LTE for UMTS - OFDMA and SC-FDMA based radio access," John Wiley & Sons, 2009.

[HZZYW10] Honglin Hu, Jian Zhang, Xiaoying Zheng, Yang Yang, and Ping Wu, "Self-configuration and self-optimization for LTE networks," *IEEE Communications Magazine*, vol.48, no.2 , pp.94-100, February 2010.

[IEEE802web07] IEEE 802.16 Broadband Wireless Access Working Group, "Multi-hop relay system evaluation methodology (channel model and performance metric)," February 2007, www.ieee802.org/16/relay/docs/80216j-06_013r3.pdf

[IEEE802web11] IEEE 802.16 Broadband Wireless Access Working Group, "Outdoor path loss models for 802.11ah" IEEE 802.11-11/0272r0," February 2011, http://www.scribd.com/doc/84339167/LTE-Path-Loss

[JBTMK10] Thomas Jansen, Irina Balan, John Turk, Ingrid Moerman, and Thomas Kurner, "Handover parameter optimization in LTE self-organizing networks," *IEEE Vehicular Technology Conference-Fall (VTC-Fall)*, pp.1–5, September 2010.

[JBW08] Peng Jiang, John Bigham, and Jiayi Wu, "Self-organizing relay stations in relay based cellular networks," *Computer Communication*, vol. 31, pp.2937-2945, Aug 2008.

[JR93a] Hua Jiang and Stephen S. Rappaport, "CBWL: A new channel assignment and sharing method for cellular communication systems," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.189-193, May 1993.

[JR93b] Hua Jiang and Stephen S. Rappaport, "Prioritized channel borrowing without locking: a channel sharing strategy for cellular communications," *IEEE Global Communications Conference (GLOBECOM)*, vol.1, pp.276-280, December 1993.

[JR94] Hua Jiang and Stephen S. Rappaport, "CBWL: a new channel assignment and sharing method for cellular communication systems," *IEEE Transactions on Vehicular Technology*, vol.43, no.2, pp.313-322, May 1994.

[JXJA08] Jun Cai, Xuemin Shen, Jon W. Mark, and Attahiru S. Alfa, "Semi-distributed user relaying algorithm for amplify-and-forward wireless relay networks,"

IEEE Transactions on wireless communications, vol.7, no.4, pp.1348-1357, April 2008.

[KAK10] Raymond Kwan, Rob Arnott, and Mitsuhiro Kubota, "On radio admission control for LTE systems," *IEEE Vehicular Technology Conference-Fall (VTC-Fall)*, pp.1-5, September 2010.

[KAPTK10] Raymond Kwan, Rob Arnott, R. Paterson, Riccardo Trivisonno, and Mitsuhiro Kubota, "On mobility load balancing for LTE systems," *IEEE Vehicular Technology Conference-Fall (VTC-Fall)*, pp.1-5, September 2010.

[KG78] Tomson Joe Kahwa and Nicolaos D. Georganas, "A hybrid channel assignment scheme in large-scale, cellular-structured mobile communication systems," *IEEE Transactions on Communications*, vol.26, no.4, pp.432-438, April 1978.

[KGSABS09] Yevgeni Koucheryavy, Giovanni Giambene, Dirk Staehle, Francisco Barcelo-Arroyo, Torsten Braun, and Vasilios Siris, "Traffic and QoS management in wireless multimedia networks: COST 290 Final Report," Springer, 2009.

[KPK08] Niko Kolehmainen, Jani Puttonen, Petteri Kela, Tapani Ristaniemi, Tero Henttonen, and Martti Moisio, "Channel quality indication reporting schemes for UTRAN Long Term Evolution downlink," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.2522-2526, May 2008.

[KSK08] Sofoklis Kyriazakos, Ioannis Soldatos, and George Karetsos, "4G mobile and wireless communications technologies," River Publishers, 2008.

[LLZL10] Weihao Lv, Wenjing Li, Heng Zhang, and Yanguang Liu, "Distributed mobility load balancing with RRM in LTE," *IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, pp.457–461, October 2010.

[LGK10] Doo-Won Lee, Gye-Tae Gil, and Dong-Hoi Kim, "A cost-based adaptive handover hysteresis scheme to minimize the handover failure rate in 3GPP LTE system," *Eurasip Journal on Wireless Communications and Networking*; 2010: Article ID 750173.

[LPS09] Byeong Gi Lee, Daeyoung Park, and Hanbyul Seo, "Wireless communications resource management," John Wiley & Sons, 2009.

[LSJB10] Andreas Lobinger, Szymon Stefanski, Thomas Jansen, and Irina Balan, "Load balancing in downlink LTE self-optimizing networks," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.1-5, May 2010.

[LW07] Jemin Lee and Hano Wang, "A multi-hop user admission algorithm for fixed relay stations with limited capabilities in OFDMA cellular networks," *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp.1-5, September 2007.

[MYYZ12] Chuan Ma, Rui Yin, Guanding Yu, and Jietao Zhang, "Reference signal power control for load balancing in downlink LTE-A self-organizing networks," *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp.460-464, September 2012.

[NADN06] Ridha Nasri, Zwi Altman, Herve Dubreil, and Zakaria Nouir, "WCDMA downlink load sharing with dynamic control of soft handover parameters," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.942-946, May, 2006.

[NA07] Ridha Nasri and Zwi Altman, "Handover adaptation for dynamic load balancing in 3GPP long term evolution systems," *International Conference on Advances in Mobile Computing and Multimedia (MoMM)*, pp.145-154, December 2007.

[NEC09] NEC, "Self organizing network: NEC's proposals for next-generation radio network management," February 2009, http://www.nec.com

[NGMN07] Next Generation Mobile Networks, "NGMN informative list of SON use cases (v1.23)," April. 2007, http://www.ngmn.org/nc/home.html

[NPPDBC03] P. Nahi, C.G. Parini, S. Papadopoulos, Lin Du, John Bigham, and Laurie Cuthbert, "A semi-smart antenna concept using real-time synthesis for use in a distributed load balancing scheme for cellular networks," *International Conference on Antennas and Propagation (ICAP)*, vol.1, pp.168-171, April. 2003.

[PB05] Christian Prehofer and Christian Bettstetter, "Self-organization in communication networks: principles and design paradigm," *IEEE Communications Magazine*, vol.43, no.7, pp.78-85, July 2005.

[PWSHYMVL04] R. Pabst, B.H. Walke, D.C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, and M. Lott, "Relay-based deployment concepts

for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol.42, no.9, pp.80-89, September 2004.

[RH12] Juan Ramiro and Khalid Hamied, "Self-organizing networks (SON): self-planning, self-optimization and self-healing for GSM, UMTS and LTE," John Wiley & Sons, 2012.

[RY10] Mahmudur Rahman and Halim Yanikomeroglu, "Enhancing cell-edge performance a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Transactions on Wireless Communications*, vol.9, no.4, pp.1414-1425, April 2010.

[SKMNT10] Mamoru Sawahashi, Yoshihisa Kishiyama, Akihito Morimoto, Datsuke Nishikawa, Motohiro Tanno, "Coordinated multipoint transmission/ reception techniques for LTE-Advanced," *IEEE Wireless Communications*, vol.17, no. 3, pp.26–34 June 2010.

[SOCRATES10] INFSO-ICT216284 FP7 SOCRATES, "Final report on self-organisation and its implications in wireless access networks," D5.9, v1.0, December 2010, http://www.fp7-socrates.org/files/Publications/

[SV09] L.C. Schmelz, J.L. van den Berg, "Self-organisation in wireless networks - use cases and their interrelations," *Wireless World Research Forum Meeting* 22, Paris, France, May 2009.

[TC11] Yuh-Ren Tsai, Cheng-Ju Chang, "Cooperative information aggregation for distributed estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol.59, no.10, pp.3876–3888. August 2011.

[TUWienLTEsimulator] Institute of Telecommunications, Technische Universität Wien, "LTE downlink system level simulator," http://www.nt.tuwien.ac.at/about-us/staff/josep-colom-ikuno/lte-simulators/

[TY03] Ozan K. Tonguz and Evsen Yanmaz, "On the theory of dynamic load balancing," *IEEE Global Communications Conference (GLOBECOM)*, vol.7, pp.3626–3630, December 2003.

[TY08] Ozan K. Tonguz and Evsen Yanmaz, "The mathematical theory of dynamic load balancing in cellular networks," *IEEE Transactions on Mobile Computing*, vol.7, no.12, pp.1504-1518, April 2008.

[WDQYT05] Hongyi Wu, Swades De, Chunming Qiao, Evsen Yanmaz, and Ozan K. Tonguz, "Handoff performance of the integrated cellular and ad hoc relaying (iCAR) system," *ACM Wireless Networks,* vol.11, no.6, pp.775-785, November 2005.

[WJL09] Liping Wang, Yusheng Ji, and Fuqiang Liu, "Resource allocation for OFDMA relay-enhanced system with cooperative selection diversity," *IEEE Wireless Communications and Networking Conference (WCNC)*, pp.1-6, April 2009.

[WTJLHL10] Xijun Wang, Hui Tian, Fan Jiang, Xiangyan Li, Xuanji Hong, and Tairi Li, "Cell-cluster based traffic load balancing in cooperative cellular networks," *IEEE Consumer Communications and Networking Conference (CCNC)*, pp.1-5, January 2010.

[WTN04] Zhenyu Wang, Eustace K. Tameh, and Andrew R. Nix, "Statistical peer-to-peer channel models for outdoor urban environments at 2GHz and 5GHz," *IEEE Vehicular Technology Conference-Fall (VTC-Fall)*, vol.7, pp.5101-5105, September, 2004.

[WZ05] Ying Wang and Ping Zhang, "Radio resource management," Beijing University of posts and telecommunication Press, May 2005.

[Xiao10] Lin Xiao, "Radio resource allocation in relay based OFDMA cellular networks," PhD Thesis, Queen Mary University of London, 2010 January.

[Yao07] Na Yao, "A CBR approach for radiation pattern control in WCDMA networks," PhD Thesis, Queen Mary University of London, January 2007.

[YLCW12] Ying Yang, Pengfei Li, Xiaohui Chen, and Weidong Wang, "A High-efficient algorithm of mobile load balancing in LTE system," *IEEE Vehicular Technology Conference-Fall (VTC-Fall)*, pp.1-5, September 2012.

[ZFLLW11] Kan Zheng, Bin Fan, Jianhua Liu, Yicheng Lin, and Wenbo Wang, "Interference coordination for OFDM-based multihop LTE-advanced networks," *IEEE Wireless Communications*, vol.18, no.1, pp.54-63, February 2011.

[ZQMZ10a] Heng Zhang, Xuesong Qiu, Luoming Meng, and Xidong Zhang, "Design of distributed and autonomic load balancing for self-organization LTE," *IEEE Vehicular Technology Conference-Spring (VTC-Spring)*, pp.1-5, Sept 2010.

[ZQMZ10b] Heng Zhang, Xuesong Qiu, Luoming Meng, and Xidong Zhang, "Achieving distributed load balancing in self-organizing LTE radio access network with autonomic network management," *IEEE Global Communications Conference (GLOBECOM)*, pp.454-459, December 2010.

[ZWLZB12] Hongtao Zhang, Xiaoxiang Wang, Yang Liu, LZ Zheng, and Thomas Michael Bohnert, "Resource allocation for relay-assisted OFDMA systems using inter-cell interference coordination," *EURASIP Journal on Wireless Communications and Networking*, Artical ID 156, January 2012.

[ZY89] Ming Zhang and Tak-Shing P. Yum, "Comparisons of channel-assignment strategies in cellular mobile telephone systems," *IEEE Transactions on Vehicular Technology*, vol.38, no.4, pp.211-215, November 1989.

[ZY91] Ming Zhang and Tak-Shing P. Yum "The nonuniform compact pattern allocation algorithm cellular mobile systems," *IEEE Transactions on Vehicular Technology*, vol.40, no.2, pp.387-391, May 1991.